Appendix: A Theory of Narrow Thinking

Chen Lian[*]

December 1, 2020

Appendix A: Proofs

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

$$E\left[u\left(y_1^*(s_1), \cdots, y_N^*(s_N), \vec{\theta}\right)\right] > E\left[u\left(x_1^*(s_1), \cdots, x_N^*(s_N), \vec{\theta}\right)\right].$$

As each $\mathcal{X}_i$ is convex, we can define

$$f(t) = E\left[u\left(x_1^*(s_1) + t\left(y_1^*(s_1) - x_1^*(s_1)\right), \cdots, x_N^*(s_1) + t\left(y_N^*(s_N) - x_N^*(s_N)\right), \vec{\theta}\right)\right].$$

From the decision-by-decision optimality condition in (3) and the fact that u is twice continuously differentiable, we have, for all i, $E\left[\frac{\partial u}{\partial x_i}\left(x_1^*(s_1), \cdots, x_N^*(s_N), \vec{\theta}\right)|s_i\right] = 0$. Moreover, we have

$$f'(0) = \sum_{i=1}^{N}\left\{E\left[\frac{\partial u}{\partial x_i}\left(x_1^*(s_1), \cdots, x_N^*(s_N), \vec{\theta}\right)\left(y_i^*(s_i) - x_i^*(s_i)\right)\right]\right\}.$$

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

$$\begin{aligned} & E\left[\frac{\partial u}{\partial x_i}\left(x_1^*(s_1), \cdots, x_N^*(s_N), \vec{\theta}\right)\left(y_i^*(s_i) - x_i^*(s_i)\right)\right] \\ = & E\left[E\left[\frac{\partial u}{\partial x_i}\left(x_1^*(s_1), \cdots, x_N^*(s_N), \vec{\theta}\right)\left(y_i^*(s_i) - x_i^*(s_i)\right)|s_i\right]\right] \\ = & E\left[E\left[\frac{\partial u}{\partial x_i}\left(x_1^*(s_1), \cdots, x_N^*(s_N), \vec{\theta}\right)|s_i\right]\left(y_i^*(s_i) - x_i^*(s_i)\right)\right] = 0. \end{aligned}$$

[*]UC Berkeley; chen_lian@berkeley.edu

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 $f(1)>f(0)=u\left(x_1^*(s_1),\cdots,x_N^*(s_N),\vec{\theta}\right)$. This is contradictory. In fact, this proposition is essentially Theorem 1 in Chapter 5 of Marschak and Radner (1972).

Comment: It is also useful to prove that the optimum of (2), if exists, is unique. If the optimum of (2) is not unique,[1] 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 $E\left[u\left(z_1^*(s_1),\cdots,z_N^*(s_N),\vec{\theta}\right)\right]>E\left[u\left(x_1^*(s_1),\cdots,x_N^*(s_N),\vec{\theta}\right)\right]$ and $E\left[u\left(z_1^*(s_1),\cdots,z_N^*(s_N),\vec{\theta}\right)\right]>E\left[u\left(y_1^*(s_1),\cdots,y_N^*(s_N),\vec{\theta}\right)\right]$. This contradicts with the optimality of $\{x_1^*(\cdot),\cdots,x_N^*(\cdot)\}$ and $\{y_1^*(\cdot),\cdots,y_N^*(\cdot)\}$.

Proof of Proposition 1. The optimality condition for each player i in the equivalent game is the same as the decision-specific optimality condition for decision i in (3). The equivalence between the Bayesian Nash Equilibrium in the Bayesian game played by multiple selves and the solution of (2) is then a direct corollary of Lemma 1.

Proof of Lemma 2, Proposition 2 and Corollary 1. Proposition 10 in Appendix D shows, in a general environment, how to derive the narrow thinker's elasticities based on her optimal decision rule. Applying it to here and from the optimal consumption decisions (6), we have: for $i\in\{1,2\}$,

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

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

$$\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{p}_i}=\frac{-\psi_i}{1-\lambda_{-i,i}\gamma_{i,-i}\gamma_{-i,i}}\quad\text{and}\quad\frac{\partial\hat{x}_{-i}^{\text{Narrow}}}{\partial\hat{p}_i}=\frac{-\psi_i\lambda_{-i,i}\gamma_{-i,i}}{1-\lambda_{-i,i}\gamma_{i,-i}\gamma_{-i,i}}.\tag{26}$$

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

$$\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{p}_i}=\omega_i\frac{\partial\hat{x}_i^{\text{Neglect}}}{\partial\hat{p}_i}+(1-\omega_i)\frac{\partial\hat{x}_i^{\text{Standard}}}{\partial\hat{p}_i},\tag{27}$$

$$\frac{\partial\hat{x}_{-i}^{\text{Narrow}}}{\partial\hat{p}_i}=\lambda_{-i,i}\left[\omega_i\frac{\partial\hat{x}_{-i}^{\text{Neglect}}}{\partial\hat{p}_i}+(1-\omega_i)\frac{\partial\hat{x}_{-i}^{\text{Standard}}}{\partial\hat{p}_i}\right],\tag{28}$$

where

$$\omega_i=\frac{1-\lambda_{-i,i}}{1-\lambda_{-i,i}\gamma_{i,-i}\gamma_{-i,i}}\in[0,1],\tag{29}$$

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

from the definition of x_i^{Neglect} in (8),

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

and

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

The comparative statics in Proposition 2 follow directly from the formula of ω_i in (9). Corollary 1 follows directly from the formula of $\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i}$ and $\frac{\partial \hat{x}_{-i}^{\text{Narrow}}}{\partial \hat{p}_i}$ in (26). Lemma 2 follows from (26) when $\sigma_{1,2}^2, \sigma_{2,1}^2 = +\infty$ ($\lambda_{1,2}, \lambda_{2,1} = 0$).

Proof of Proposition 3. Taking an unconditional expectation of (6) averaging over the realization of all fundamentals and signals, we have

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

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

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

We then have

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

Proof of Lemma 3. Under narrow thinking, if condition (17) holds, averaging over the realization of noises in signals, we then have

$$-\kappa_i \hat{x}_i^{\text{Narrow}}\left(\hat{p}_1, \cdots, \hat{p}_N\right) - \hat{p}_i = -\kappa_j \hat{x}_j^{\text{Narrow}}\left(\hat{p}_1, \cdots, \hat{p}_N\right) - \hat{p}_j,$$

and

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

This is inconsistent with the formula about the narrow thinker's demand elasticities in the proof of Proposition 4.

Proof of Proposition 4 and Corollary 3. Proposition 10 in Appendix D shows, in a general environment, how to derive the narrow thinker's elasticities based on her optimal decision rule. Applying it to here and from conditions (14) and (15), we have, for all i,

$$
\begin{aligned}
-\kappa_i \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} &= 1 - \kappa_y \frac{\partial \hat{y}^{\text{Narrow}}}{\partial \hat{p}_i}, \\
&= 1 + \frac{\kappa_y}{\mu_y} \left(\sum_{j=1}^{N} \mu_j \frac{\partial \hat{x}_j^{\text{Narrow}}}{\partial \hat{p}_i} + \mu_i \right),
\end{aligned}
$$

and for all $k \neq i$,

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

Together, we have, for all i,

$$
\begin{aligned}
\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} &= -1 - \frac{1-\kappa_i}{\kappa_i} \left(\frac{\sum_{j \neq i} \frac{\mu_j}{\kappa_j} \frac{\frac{\mu_y}{\kappa_y} \lambda_{j,i}}{\frac{\mu_y}{\kappa_y} + (1-\lambda_{j,i}) \frac{\mu_j}{\kappa_j}} + \frac{\mu_y}{\kappa_y}}{\frac{\mu_i}{\kappa_i} + \sum_{j \neq i} \frac{\mu_j}{\kappa_j} \frac{\frac{\mu_y}{\kappa_y} \lambda_{j,i}}{\frac{\mu_y}{\kappa_y} + (1-\lambda_{j,i}) \frac{\mu_j}{\kappa_j}} + \frac{\mu_y}{\kappa_y}} \right), \\
\frac{\partial \hat{y}^{\text{Narrow}}}{\partial \hat{p}_i} &= \frac{\mu_i \frac{1-\kappa_i}{\kappa_i}}{\kappa_y \left(\frac{\mu_i}{\kappa_i} + \sum_{j \neq i} \frac{\mu_j}{\kappa_j} \frac{\frac{\mu_y}{\kappa_y} \lambda_{j,i}}{\frac{\mu_y}{\kappa_y} + (1-\lambda_{j,i}) \frac{\mu_j}{\kappa_j}} + \frac{\mu_y}{\kappa_y} \right)}.
\end{aligned} \tag{30}
$$

And for all $k \neq i$,

$$
\frac{\partial \hat{x}_k^{\text{Narrow}}}{\partial \hat{p}_i} = \frac{1}{\kappa_k} \frac{\lambda_{k,i} \frac{\mu_y}{\kappa_y}}{\frac{\mu_y}{\kappa_y} + (1-\lambda_{k,i}) \frac{\mu_k}{\kappa_k}} \frac{\mu_i \frac{1-\kappa_i}{\kappa_i}}{\left(\frac{\mu_i}{\kappa_i} + \sum_{j \neq i} \frac{\mu_j}{\kappa_j} \frac{\frac{\mu_y}{\kappa_y} \lambda_{j,i}}{\frac{\mu_y}{\kappa_y} + (1-\lambda_{j,i}) \frac{\mu_j}{\kappa_j}} + \frac{\mu_y}{\kappa_y} \right)}. \tag{31}
$$

From (30), we have, for $i \in \{1, \cdots, N\}$,

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

where

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

$\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_i} = -1 - \frac{1-\kappa_i}{\kappa_i} \left(\frac{\sum_{j \neq i} \frac{\mu_j}{\kappa_j} + \frac{\mu_y}{\kappa_y}}{\frac{\mu_i}{\kappa_i} + \sum_{j \neq i} \frac{\mu_j}{\kappa_j} + \frac{\mu_y}{\kappa_y}} \right)$, and $\frac{\partial \hat{x}_i^{\text{Explicit}}}{\partial \hat{p}_i} = -1$ based on (18). This proves Proposition 4.

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

3 then follows from Proposition 4.

Proof of Proposition 5. First define $f(a,b,c,x) \equiv \frac{(a+x)(b+c+x)}{(a+c+x)(b+x)}$, where $a,b,c,x>0$ and $b>a$. Based on the formula for ω_i in (21), we have $\omega_i = 1 - f(a,b,c,x)$ where $a = \sum_{j\neq i} \frac{\mu_j}{\kappa_j} \frac{\frac{\mu_y}{\kappa_y}\lambda_{j,i}}{\frac{\mu_y}{\kappa_y}+(1-\lambda_{j,i})\frac{\mu_j}{\kappa_j}}$, $b = \sum_{j\neq i}\frac{\mu_j}{\kappa_j}$, $c = \frac{\mu_i}{\kappa_i}$ and $x = \frac{\mu_y}{\kappa_y}$. Note that $b>a$ because $\frac{\mu_y}{\kappa_y} = \frac{\frac{\bar{y}}{w}}{-\frac{h''(\bar{y})\bar{y}}{h'(\bar{y})}} = -\frac{h'(\bar{y})}{wh''(\bar{y})}$ is always positive even if $\bar{y}$ is negative.

Let me prove (iii) first. For this to be true, I only need to show that $f(a,b,c,x)$ increases in a, which is true because $b>a$.

To prove (i), I only need to show that $f(a,b,c,x)$ increases in both a and x. The latter is true because

$$
\begin{aligned}
\frac{\partial f(a,b,c,x)}{\partial x} &= \frac{(a+b+c)+2x}{(b+x)(a+c+x)} - \frac{[(a+b+c)+2x](a+x)(b+c+x)}{(b+x)^2(a+c+x)^2} \\
&= \frac{(a+b+c)+2x}{(b+x)(a+c+x)}\left(1-\frac{(a+x)(b+c+x)}{(b+x)(a+c+x)}\right) > 0.
\end{aligned}
$$

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

To prove (iv), we start from (21) and take the limit of $\lambda_{j,i}\to 0$ for all $j\neq i$ and $\frac{\mu_y}{\kappa_y}\to 0$.

To prove (v), we start from (21) and take the limit of $\mu_i/\kappa_i \to 0$.

Proof of Corollary 2. When, for all i and $j\neq i\in\{1,\cdots,N\}$, $\kappa_i=\kappa_y=\kappa$ and $\lambda_{j,i}=\lambda$, the formula for ω_i in (21) becomes

$$
\begin{aligned}
\omega_i &= 1 - \frac{\left(\sum_{j\neq i}\frac{\mu_j}{\mu_y}\frac{\mu_y\lambda}{\mu_y+(1-\lambda)\mu_j}+1\right)\left(\frac{\mu_x}{\mu_y}+1\right)}{\left(\sum_{j\neq i}\frac{\mu_j}{\mu_y}\frac{\lambda\mu_y}{\mu_y+(1-\lambda)\mu_j}+\frac{\mu_i}{\mu_y}+1\right)\left(\frac{\mu_x}{\mu_y}\left(1-\frac{\mu_i}{\mu_x}\right)+1\right)} \\
&\geq 1 - \frac{\left(\sum_{j\neq i}\frac{\mu_j}{\mu_y}\lambda+1\right)\left(\frac{\mu_x}{\mu_y}+1\right)}{\left(\sum_{j\neq i}\frac{\mu_j}{\mu_y}\lambda+\frac{\mu_i}{\mu_y}+1\right)\left(\frac{\mu_x}{\mu_y}\left(1-\frac{\mu_i}{\mu_x}\right)+1\right)} \\
&= 1 - \frac{\left(\lambda\frac{\mu_x}{\mu_y}\left(1-\frac{\mu_i}{\mu_x}\right)+1\right)\left(\frac{\mu_x}{\mu_y}+1\right)}{\left((1-\lambda)\frac{\mu_i}{\mu_x}\frac{\mu_x}{\mu_y}+\lambda\frac{\mu_x}{\mu_y}+1\right)\left(\frac{\mu_x}{\mu_y}\left(1-\frac{\mu_i}{\mu_x}\right)+1\right)} = \underline{\omega}_i,
\end{aligned}
$$

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

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

$$
\hat{\varphi}_i - \kappa_i \hat{x}_i^*(s_i) = -\kappa_y E_i[\hat{y}^*] \qquad \forall i, \tag{32}
$$

$$
\sum_{i=1}^N \mu_i \hat{x}_i^*(s_i) + \mu_y \hat{y}^* = 0, \tag{33}
$$

where, as in Section 4.1, $\kappa_y = -\frac{h''(\bar{y})\bar{y}}{h'(\bar{y})}$, $\mu_i = \frac{\bar{p}_i \bar{x}_i}{w}$, and $\mu_y = \frac{\bar{y}}{w}$. Based on those two conditions and applying Proposition 10, we have, for all i and $k \neq i$,

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

$$-\kappa_k \frac{\partial \hat{x}_k^{\text{Narrow}}}{\partial \hat{\varphi}_i} = \frac{\kappa_y}{\mu_y} \left[\mu_k \frac{\partial \hat{x}_k^{\text{Narrow}}}{\partial \hat{\varphi}_i} + \lambda_{k,i} \sum_{j \neq k} \mu_j \frac{\partial \hat{x}_j^{\text{Narrow}}}{\partial \hat{\varphi}_i} \right],$$

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

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

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

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

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

where the weight $\omega_i \in [0, 1]$ is still given by (21), $\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{\varphi}_i} = \frac{1}{\kappa_i} \frac{\sum_{j \neq i} \frac{\mu_j}{\kappa_j} + \frac{\mu_y}{\kappa_y}}{\frac{\mu_i}{\kappa_i} + \sum_{j \neq i} \frac{\mu_j}{\kappa_j} + \frac{\mu_y}{\kappa_y}} > 0$, and $\frac{\partial \hat{x}_i^{\text{Explicit}}}{\partial \hat{\varphi}_i} = 0$ based on (18).

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

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

$$-\kappa_i \hat{x}_i^*(s_i) = -\kappa_y E_i[\hat{y}^*], \tag{34}$$

$$\sum_{i=1}^{N} \mu_i \hat{x}_i^*(s_i) + \mu_y \hat{y}^* = \sum_{i=1}^{N} \mu_i^w \hat{w}_i, \tag{35}$$

where $\kappa_y = -\frac{h''(\bar{y})\bar{y}}{h'(\bar{y})}$, $\mu_i = \frac{\bar{p}_i \bar{x}_i}{w}$, $\mu_y = \frac{\bar{y}}{w}$, and $\mu_i^w = \frac{\bar{w}_i}{w}$. Based on those two conditions and applying Proposition 10, we have, for all i and $k \neq i$,

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

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

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

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

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

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

where the weight $\omega_i \in [0, 1]$ is still given by (21), $\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{w}_i} = \frac{\kappa_y}{\kappa_i} \frac{\mu_i^w}{\kappa_y \left(\frac{\mu_i}{\kappa_i} + \sum_{j \neq i} \frac{\mu_j}{\kappa_j} + \frac{\mu_y}{\kappa_y} \right)} > 0$, and $\frac{\partial \hat{x}_i^{\text{Explicit}}}{\partial \hat{w}_i} = \frac{\mu_i^w}{\mu_i}$. Here, by explicit budgeting, I mean the consumer uses the entirety of w_i on the consumption of good i.

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

Appendix B: Consumer Theory

Demand Elasticities in Response to Temporary Shocks and Persistent Differences

As discussed after Proposition 3, the narrow thinker's demand elasticity estimated based on temporary price shocks can differ from the one estimated based on persistent price differences. Let me expand here.

Consider a sample of otherwise identical consumers. They all have the same utility as in Section 3. Each consumer solves the same consumer problem but may face different prices. Same as Section 3, each self knows the price of the good she buys, but only receives noisy signals about other prices faced by her other selves. All consumers share the same signal-to-noise ratio of their signals (thus the same $\{\lambda_{i,j}\}$). As different consumers have the same utility and the same $\{\lambda_{i,j}\}$, they all share the same demand elasticities.

Let us now consider two cases. In the first case, different consumers face the same distribution of prices but may face different price realizations. That is, each consumer faces an independent draw from the same price distribution. By studying how otherwise identical consumers respond to different, temporary, price realizations, we can identify the narrow thinker's demand elasticities: $\left\{\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_j}\right\}$.

In the second case, different consumers may face different price distributions. In particular, the mean of the price distributions faced by different consumers may be different. Instead of studying responses to temporary price shocks, we now study how the mean of each consumer's average demand (across time) varies with the mean of the price she faces. As illustrated in Proposition 3, the narrow thinker is unbiased on average. This approach will then identify demand elasticities under standard consumer theory: $\left\{\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_j}\right\}$.

Based on the demand elasticities estimated from these two cases, one can then recover the degree of narrow thinking λ. For example, consider the environment in Section 3 and assume homogeneity between the two goods $i \in \{1, 2\}$. We have, from the proof of Proposition 2,

$$\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} = \frac{-\psi}{1 - \lambda\gamma^2} \quad \text{and} \quad \frac{\partial \hat{x}_{-i}^{\text{Narrow}}}{\partial \hat{p}_i} = \frac{-\psi\lambda\gamma}{1 - \lambda\gamma^2}$$

and

$$\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_i} = \frac{-\psi}{1 - \gamma^2} \quad \text{and} \quad \frac{\partial \hat{x}_{-i}^{\text{Standard}}}{\partial \hat{p}_i} = \frac{-\psi\gamma}{1 - \gamma^2}.$$

The degree of narrow thinking λ is then given by:

$$\lambda = \left(\frac{\partial \hat{x}_{-i}^{\text{Narrow}}}{\partial \hat{p}_i} / \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i}\right) / \left(\frac{\partial \hat{x}_{-i}^{\text{Standard}}}{\partial \hat{p}_i} / \frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_i}\right).$$

Observational Equivalence with Other Forms of Bounded Recall.

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

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

$$\hat{x}_i^*(s_i) = E_i\left[-\psi\hat{p}_i + \gamma\hat{x}_{-i}^*(s_{-i})\right], \tag{36}$$

Each self $i \in \{1, 2\}$ of the narrow thinker, who is in charge of purchasing good i, perfectly knows $p_i \sim \log\mathcal{N}\left(\log\bar{p}, \sigma_p^2\right)$, but receives a noisy signal about each of the other p_{-i}: $s_{i,-i} = p_{-i}\epsilon_{i,-i}$, with $\epsilon_{i,-i} \sim \log\mathcal{N}\left(0, \sigma^2\right)$ and $\sigma^2 > 0$. All ϵ and p are independent of each other. In other words, for $i \in \{1, 2\}$, $s_i = \{p_i, s_{i,-i}\}$.

Now, consider a different decision maker, whose bounded recall is captured by a noisy signal about the past endogenous decision. For such a decision maker (BR), in the case that decision 1 is made before decision 2, her self 1's information is given by $s_1 = \left\{p_1, s_{1,2}^{\text{BR}} = p_2\epsilon_{1,2}^{\text{BR}}\right\}$, where $\epsilon_{1,2}^{\text{BR}} \sim \log\mathcal{N}\left(0, \sigma_{\text{BR}}^2\right)$. That is, she perfectly knows p_1, receives a noisy signal about the future p_2, and σ_{BR}^2 captures the size of the noise. On the other hand, her self 2's information is given by $s_2 = \left\{p_2, s_{2,1}^{\text{BR}} = x_1\epsilon_{2,1}^{\text{BR}}\right\}$, where $\epsilon_{2,1}^{\text{BR}} \sim \log\mathcal{N}\left(0, \Sigma_{\text{BR}}^2\right)$. That is, she perfectly knows p_2, but cannot perfectly recall her past decision x_1. Instead, she receives a noisy signal about x_1 and Σ_{BR}^2 captures the size of the noise. Similarly, for the case that the decision 2 is made before the decision 1, we have $s_2 = \left\{p_2, s_{2,1}^{\text{BR}} = p_1\epsilon_{2,1}^{\text{BR}}\right\}$ and $s_1 = \left\{p_1, s_{1,2}^{\text{BR}} = x_2\epsilon_{1,2}^{\text{BR}}\right\}$, where $\epsilon_{2,1}^{\text{BR}} \sim \log\mathcal{N}\left(0, \sigma_{\text{BR}}^2\right)$ and $\epsilon_{1,2}^{\text{BR}} \sim \log\mathcal{N}\left(0, \Sigma_{\text{BR}}^2\right)$.

Note that the information and behavior of such a decision maker (BR), whose bounded recall is captured by a noisy signal about the past endogenous decision, will depend on the order of the decisions. To compare her behavior with that of the above narrow thinker, we define $\hat{x}_i^{\text{BR}}\left(\hat{p}_1, \hat{p}_2\right) \equiv E\left[\hat{x}_i^*\left(s_i\right)|\hat{p}_1, \hat{p}_2\right]$, where $E\left[\cdot|\hat{p}_1, \hat{p}_2\right]$ averages over not only the realization of noises in signals but also the potential order of decisions.[2]

In fact, without additional knowledge about the order of decisions, the own-price and cross-price elasticities of the narrow thinker and those of the decision maker (BR) are observationally equivalent.[3]

Lemma 4 *For a narrow thinker with any size of noise $\sigma^2 > 0$, one can find a decision maker (BR) with sizes of noise $\sigma_{BR}^2 > 0$ and $\Sigma_{BR}^2 > 0$, such that the two decision makers share the same own-price and cross-price elasticities:*

$$\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} = \frac{\partial \hat{x}_i^{\text{BR}}}{\partial \hat{p}_i} \quad \textit{and} \quad \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_{-i}} = \frac{\partial \hat{x}_i^{\text{BR}}}{\partial \hat{p}_{-i}} \quad \forall i \in \{1, 2\}. \tag{37}$$

This equivalence result should not be surprising: in the end, the noisy signals in both cases capture bounded recall and restrict the later decision to be made based on an imperfect perception of the earlier decision. One may then wonder why I focus on the case with noisy signals about fundamentals throughout the paper, instead of the case with noisy signals directly about decisions. As mentioned above, one advantage of the former case is that the analysis does not require knowledge about the exact order of decisions. Moreover, for the case with noisy signals directly about decisions, it is hard to analytically characterize the decision maker's behavior when we have $N \geq 3$ decisions. To illustrate the difficulty, consider self N, who receives noisy signals about the first $N-1$ decisions: $s_{N,i}^{BR} = x_i\epsilon_{N,i}^{BR}, i \in \{1, \cdots, N-1\}$. As each decision is a function of all fundamentals $(p_1, \cdots, p_N)$, the signal about each decision i will also be informative about all other decisions and vice versa. In other words, the "rational confusion" among signals is prevalent, and the analysis becomes intractable. On the other hand, with noisy signals about (independent) fundamentals, one can still characterize analytically the narrow thinker's behavior in many interesting economic environments,

[2] I assume with 50% of probability that decision 1 is made before decision 2, and with another 50% of probability that decision 2 is made before decision 1.

[3] Such a decision maker (BR) is still a narrow thinker based on the general definition in Definition 1, but has different information from the narrow thinker studied in Section 3 and beyond.

as studied in Section 4.

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

Proof of Lemma 4. Under narrow thinking, from the proof of Proposition 2, we have, for $i \in \{1, 2\}$,

$$\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} = -\psi \left(1 + \frac{\lambda\gamma^2}{1 - \lambda\gamma^2}\right) \quad \text{and} \quad \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_{-i}} = \frac{-\psi\lambda\gamma}{1 - \lambda\gamma^2}, \tag{38}$$

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

$$\begin{aligned}
\hat{x}_1^* (s_1) &= \alpha_1 \hat{p}_1 + \alpha_2 \hat{s}_{1,2}^{\text{BR}} = \alpha_1 \hat{p}_1 + \alpha_2 \left(\hat{p}_2 + \hat{\epsilon}_{1,2}^{\text{BR}}\right), \\
\hat{x}_2^* (s_2) &= \beta_2 \hat{p}_2 + \beta_1 \hat{s}_{2,1}^{\text{BR}} = \beta_2 \hat{p}_2 + \beta_1 \left(\hat{x}_1^* (s_1) + \hat{\epsilon}_{2,1}^{\text{BR}}\right).
\end{aligned}$$

From self 1's optimality, we then have

$$\hat{x}_1^* (s_1) = -\psi \hat{p}_1 + E_1 \left[\gamma \hat{x}_2^* (s_2)\right] = -\psi \hat{p}_1 + \gamma E_1 \left[\beta_2 \hat{p}_2 + \beta_1 \hat{x}_1^* (s_1)\right].$$

As a result,

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

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

$$\begin{aligned}
\hat{x}_2^* (s_2) &= -\psi \hat{p}_2 + \gamma E_2 \left[\alpha_1 \hat{p}_1 + \alpha_2 \left(\hat{p}_2 + \hat{\epsilon}_{1,2}^{\text{BR}}\right)\right], \\
&= (-\psi + \gamma \alpha_2) \hat{p}_2 + \gamma E_2 \left[\alpha_1 \hat{p}_1 + \alpha_2 \hat{\epsilon}_{1,2}^{\text{BR}}\right], \\
&= (-\psi + \gamma \alpha_2) \hat{p}_2 + \lambda_2^{\text{BR}} \gamma \left(\hat{x}_1^* (s_1) + \hat{\epsilon}_{2,1}^{\text{BR}} - \alpha_2 \hat{p}_2\right),
\end{aligned}$$

where $\lambda_2^{\text{BR}} = \frac{\alpha_1^2 Var(p_1) + \alpha_2^2 \sigma_{\text{BR}}^2}{\alpha_1^2 Var(p_1) + \alpha_2^2 \sigma_{\text{BR}}^2 + \Sigma_{\text{BR}}^2}$. As a result,

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

Together, we have

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

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

$$\hat{x}_2^*(s_2) = \alpha_1\hat{p}_2 + \alpha_2\left(\hat{p}_1 + \hat{\epsilon}_{2,1}^{\mathrm{BR}}\right),$$
$$\hat{x}_1^*(s_1) = \beta_2\hat{p}_1 + \beta_1\left(\hat{x}_2^*(s_2) + \hat{\epsilon}_{1,2}^{\mathrm{BR}}\right).$$

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

$$\frac{\partial\hat{x}_i^{\mathrm{BR}}}{\partial\hat{p}_i} = \frac{1}{2}\left(\alpha_1+\beta_2+\beta_1\alpha_2\right) = -\psi\left[1+\frac{\gamma}{2}\left(\frac{\lambda_2^{\mathrm{BR}}\gamma}{1-\lambda_2^{\mathrm{BR}}\gamma^2}+\frac{\lambda_1^{\mathrm{BR}}\gamma}{1-\gamma^2\left(\lambda_2^{\mathrm{BR}}\left(1-\lambda_1^{\mathrm{BR}}\right)+\lambda_1^{\mathrm{BR}}\right)}\right)\right] \tag{39}$$
$$\frac{\partial\hat{x}_i^{\mathrm{BR}}}{\partial\hat{p}_{-i}} = \frac{1}{2}\left(\alpha_2+\beta_1\alpha_1\right) = -\frac{\psi}{2}\left(\frac{\lambda_2^{\mathrm{BR}}\gamma}{1-\lambda_2^{\mathrm{BR}}\gamma^2}+\frac{\lambda_1^{\mathrm{BR}}\gamma}{1-\gamma^2\left(\lambda_2^{\mathrm{BR}}\left(1-\lambda_1^{\mathrm{BR}}\right)+\lambda_1^{\mathrm{BR}}\right)}\right). \tag{40}$$

Compared to the formula about own- and cross- sensitivities under narrow thinking in (38), to prove Lemma 4, we only need to prove that, for any $\lambda \in (0,1)$, we can find $\sigma_{BR}^2, \Sigma_{BR}^2 > 0$ such that

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

This is true as

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

and

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

A General Consumer Theory Problem

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

The decision maker's utility depends on her consumption of N goods, $(x_1, \cdots, x_N)$, and the numeraire

$y \in \mathbb{R}$ (which can be interpreted as saving/borrowing or money). Her utility is given by

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

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

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

where p_i is good i's price and w is the decision maker's total wealth (treated as a constant, as I am interested in response to price shocks here).[4]

I let $\tilde{u}$ be well defined for all $y \in \mathbb{R}$. Based on the discussion about constraint problems in Section 2, this allows the "residual decision" y to be negative and guarantees that the budget constraint, $\sum_{i=1}^{N} p_i x_i + y \leq w$, will always be satisfied. As the budget constraint always binds in the optimum, one can use it to substitute y:

$$u(x_1, \cdots, x_N, \vec{p}) = \tilde{u}\left(x_1, \cdots, x_N, w - \sum_{i=1}^{N} p_i x_i\right). \tag{41}$$

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

Appendix C: A Smooth Model of Mental Accounting

Each Self i is in Charge of a Group of Decisions.

In the main analysis, I let each self i be in charge of consumption of good i, x_i. An alternative is that each self i is in charge of a group of decisions $(x_{i,1}, \cdots, x_{i,M_i})$, e.g., all goods in a spending category. I use x_i to capture the composite consumption for this group and p_i to denote the price index for this group:

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

where $\mathcal{V}_i$ is homogenous of degree one. Self i perfectly knows the price of all goods $(p_{i,1}, \cdots, p_{i,M_i})$ and receives a noisy signal about the price index of other groups: $s_{i,k} = p_k \epsilon_{i,k}$, for $k \neq i$.

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

[4]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.

By the same token, one can interpret Corollary 3 as how the composite consumption of a group/spending-category decreases excessively in response to a common price shock to all goods within that group/spending-category.

Main Results under Linearization.

Here, I use the main mental accounting application in Section 4.1 to illustrate that one can establish parallel results with linearization instead of log-linearization. In fact, the weight on mental accounting in Propositions 4 and 5 will remain to be exactly the same. The exercise here also clarifies that the limit result in part (iv) of Proposition 5 has nothing to do with log-linearization.

Specifically, consider the environment in Section 4.1. Now, I work with the linearization now. I use a tilde over a variable to denote its deviation from the point where each price is fixed at $\bar{p}_i$ and each decision is made with perfect knowledge of all prices: $\{\bar{x}_i\}_{i=1}^N = \arg\max_{\{x_i\}_{i=1}^N} u(x_1, \cdots, x_N, \bar{p}_1, \cdots \bar{p}_N)$.[5] For analytical solutions, fundamentals and noises are now normally distributed, with $p_i \sim \mathcal{N}\left(\bar{p}_i, \sigma_{p_i}^2\right)$ and $s_{i,k} = p_k + \epsilon_{i,k}$, where $\epsilon_{i,k} \sim \mathcal{N}\left(0, \sigma_{i,k}^2\right)$. All ϵ and p are independent of each other. That is, for self $i \in \{1, \cdots, N\}$, her information is given by $s_i = \{p_i, s_{i,k}\}_{k \neq i,\ k \in \{1, \cdots, N\}}$.

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

$$-\kappa_i \frac{\tilde{x}_i^*(s_i)}{\bar{x}_i} = \frac{\tilde{p}_i}{\bar{p}_i} + \frac{h''(\bar{y})}{h'(\bar{y})} E_i[\tilde{y}^*], \tag{42}$$

where the budget constraint is given by

$$\sum_{i=1}^N (\bar{p}_i \tilde{x}_i^*(s_i) + \tilde{p}_i \bar{x}_i) + \tilde{y}^* = 0. \tag{43}$$

Similar to the main analysis, I define the narrow thinker's demand function as $\tilde{x}_i^{\text{Narrow}}(\tilde{p}_1, \cdots, \tilde{p}_N) \equiv E[\tilde{x}_i^*(s_i) | \tilde{p}_1, \cdots, \tilde{p}_N]$, averaging over the realization of noises in signals. I can now re-establish Proposition 4.

Proposition 8 *The narrow thinker's own-price demand gradient is given by: for $i \in \{1, \cdots, N\}$,*

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

where ω_i shares the exact same formula as in (21) and $\frac{\partial \tilde{x}_i^{Explicit}}{\partial \tilde{p}_i} \equiv -\frac{\bar{x}_i}{\bar{p}_i}$ comes from the linearization of the explicit mental budget model in (18).

Proposition 8 shows that the narrow thinker's own-price demand gradient is given by a weighted average between that in standard consumer theory and that with explicit budgeting. Moreover, the weight ω_i

[5] Specifically, we have $\tilde{x}_i = x_i - \bar{x}_i$ and $\tilde{y} = y - \bar{y}$, where $\bar{y} = w - \sum_{i=1}^N \bar{p}_i \bar{x}_i$.

shares the exact same formula as in (21). This should not come at a surprise, because, at the point of linearization/log-linearization, the demand gradient is just a multiple of the demand elasticity in Proposition 4: $\frac{\partial \tilde{x}_i^{\text{Narrow}}}{\partial \tilde{p}_i} = \frac{\bar{x}_i}{\bar{p}_i}\frac{\partial \hat{x}_i^{\text{Explicit}}}{\partial \hat{p}_i}$.

Proof of Proposition 8. From conditions (42) and (43), and similar to the proof of Proposition 10, we have, for all i,

$$\begin{aligned}
-\frac{\kappa_i}{\bar{x}_i}\frac{\partial \tilde{x}_i^{\text{Narrow}}}{\partial \tilde{p}_i} &= \frac{1}{\bar{p}_i} + \frac{h''(\bar{y})}{h'(\bar{y})}\frac{\partial \tilde{y}^{\text{Narrow}}}{\partial \tilde{p}_i}, \\
&= \frac{1}{\bar{p}_i} - \frac{h''(\bar{y})}{h'(\bar{y})}\left(\sum_{j=1}^{N}\bar{p}_j\frac{\partial \tilde{x}_j^{\text{Narrow}}}{\partial \tilde{p}_i} + \bar{x}_i\right),
\end{aligned}$$

and for all $k \neq i$,

$$\begin{aligned}
-\frac{\kappa_k}{\bar{x}_k}\frac{\partial \tilde{x}_k^{\text{Narrow}}}{\partial \tilde{p}_i} &= \frac{h''(\bar{y})}{h'(\bar{y})}\lambda_{k,i}\frac{\partial \tilde{y}^{\text{Narrow}}}{\partial \tilde{p}_i} - \frac{h''(\bar{y})}{h'(\bar{y})}(1-\lambda_{k,i})\,\bar{p}_k\frac{\partial \tilde{x}_k^{\text{Narrow}}}{\partial \tilde{p}_i}, \\
&= -\frac{h''(\bar{y})}{h'(\bar{y})}\left[\bar{x}_k\frac{\partial \tilde{x}_k^{\text{Narrow}}}{\partial \tilde{p}_i} + \lambda_{k,i}\left(\sum_{j\neq k}\bar{p}_j\frac{\partial \tilde{x}_j^{\text{Narrow}}}{\partial \tilde{p}_i} + \bar{x}_i\right)\right].
\end{aligned}$$

Together, we have, for all i,

$$\begin{aligned}
\frac{\partial \tilde{x}_i^{\text{Narrow}}}{\partial \tilde{p}_i} &= \frac{\bar{x}_i}{\bar{p}_i}\left(-1-\frac{1-\kappa_i}{\kappa_i}\left(\frac{\sum_{j\neq i}\frac{\bar{p}_j\bar{x}_j}{\kappa_j}\frac{-\frac{h'(\bar{y})}{h''(\bar{y})}\lambda_{j,i}}{-\frac{h'(\bar{y})}{h''(\bar{y})}+(1-\lambda_{j,i})\frac{\bar{p}_j\bar{x}_j}{\kappa_j}} - \frac{h'(\bar{y})}{h''(\bar{y})}}{\frac{\bar{p}_i\bar{x}_i}{\kappa_i}+\sum_{j\neq i}\frac{\bar{p}_j\bar{x}_j}{\kappa_j}\frac{-\lambda_{j,i}\frac{h'(\bar{y})}{h''(\bar{y})}}{-\frac{h'(\bar{y})}{h''(\bar{y})}+(1-\lambda_{j,i})\frac{\bar{p}_j\bar{x}_j}{\kappa_j}} - \frac{h'(\bar{y})}{h''(\bar{y})}}\right)\right), \qquad (44)\\
\frac{\partial \tilde{y}^{\text{Narrow}}}{\partial \tilde{p}_i} &= \frac{\bar{x}_i\frac{1-\kappa_i}{\kappa_i}}{-\frac{h''(\bar{y})}{h'(\bar{y})}\left(\frac{\bar{p}_i\bar{x}_i}{\kappa_i}+\sum_{j\neq i}\frac{\bar{p}_j\bar{x}_j}{\kappa_j}\frac{-\lambda_{j,i}\frac{h'(\bar{y})}{h''(\bar{y})}}{-\frac{h'(\bar{y})}{h''(\bar{y})}+(1-\lambda_{j,i})\frac{\bar{p}_j\bar{x}_j}{\kappa_j}} - \frac{h'(\bar{y})}{h''(\bar{y})}\right)}.
\end{aligned}$$

And for all $k \neq i$,

$$\frac{\partial \tilde{x}_k^{\text{Narrow}}}{\partial \tilde{p}_i} = \frac{\bar{x}_k}{\kappa_k}\left[\frac{-\lambda_{k,i}\frac{h'(\bar{y})}{h''(\bar{y})}}{-\frac{h'(\bar{y})}{h''(\bar{y})}+(1-\lambda_{k,i})\frac{\bar{p}_k\bar{x}_k}{\kappa_k}}\frac{\bar{x}_i\frac{1-\kappa_i}{\kappa_i}}{\frac{\bar{p}_i\bar{x}_i}{\kappa_i}+\sum_{j\neq i}\frac{\bar{p}_j\bar{x}_j}{\kappa_j}\frac{-\lambda_{j,i}\frac{h'(\bar{y})}{h''(\bar{y})}}{-\frac{h'(\bar{y})}{h''(\bar{y})}+(1-\lambda_{j,i})\frac{\bar{p}_j\bar{x}_j}{\kappa_j}} - \frac{h'(\bar{y})}{h''(\bar{y})}}\right]. \qquad (45)$$

From (44), we have, for $i \in \{1, \cdots, N\}$,

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

where

$$\omega_i = 1 - \frac{\left(\sum_{j\neq i} \frac{\bar{p}_j\bar{x}_j}{\kappa_j} \frac{-\frac{h'(\bar{y})}{h''(\bar{y})}\lambda_{j,i}}{-\frac{h'(\bar{y})}{h''(\bar{y})}+(1-\lambda_{j,i})\frac{\bar{p}_j\bar{x}_j}{\kappa_j}} - \frac{h'(\bar{y})}{h''(\bar{y})}\right)\left(\sum_{j\neq i}\frac{\bar{p}_j\bar{x}_j}{\kappa_j} + \frac{\bar{p}_i\bar{x}_i}{\kappa_i} - \frac{h'(\bar{y})}{h''(\bar{y})}\right)}{\left(\frac{\bar{p}_i\bar{x}_i}{\kappa_i} + \sum_{j\neq i}\frac{\bar{p}_j\bar{x}_j}{\kappa_j}\frac{-\lambda_{j,i}\frac{h'(\bar{y})}{h''(\bar{y})}}{-\frac{h'(\bar{y})}{h''(\bar{y})}+(1-\lambda_{j,i})\frac{\bar{p}_j\bar{x}_j}{\kappa_j}} - \frac{h'(\bar{y})}{h''(\bar{y})}\right)\left(\sum_{j\neq k}\frac{\bar{p}_j\bar{x}_j}{\kappa_j} - \frac{h'(\bar{y})}{h''(\bar{y})}\right)}, \tag{46}$$

$\frac{\partial \tilde{x}_i^{\text{Standard}}}{\partial \tilde{p}_i} = \frac{\bar{x}_i}{\bar{p}_i}\left(-1 - \frac{1-\kappa_i}{\kappa_i}\left(\frac{\sum_{j\neq i}\frac{\bar{p}_j\bar{x}_j}{\kappa_j} - \frac{h'(\bar{y})}{h''(\bar{y})}}{\frac{\bar{p}_i\bar{x}_i}{\kappa_i} + \sum_{j\neq i}\frac{\bar{p}_j\bar{x}_j}{\kappa_j} - \frac{h'(\bar{y})}{h''(\bar{y})}}\right)\right)$, and $\frac{\partial \tilde{x}_i^{\text{Explicit}}}{\partial \tilde{p}_i} = -\frac{\bar{x}_i}{\bar{p}_i}$ based on (18). Now, using the fact that $\frac{\mu_y}{\kappa_y} = \frac{\frac{\bar{y}}{w}}{-\frac{h''(\bar{y})\bar{y}}{h'(\bar{y})}} = -\frac{1}{w}\frac{h'(\bar{y})}{h''(\bar{y})}$ and $\mu_i = \frac{\bar{p}_i\bar{x}_i}{w}$ for all i, we can re-write ω_i in (46) as

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

which is the exact same formula of ω_i in (21) of Proposition 4.

Because of the shared formula, Proposition 5 also follows directly.

Comment: Let me now further clarify the limit result in part (iv) of Proposition 5. Rigorously, (42) and (43) can be written as

$$-\kappa_i\frac{\tilde{x}_i^*(s_i)}{\bar{x}_i} = \frac{\tilde{p}_i}{\bar{p}_i} + \frac{h''(\bar{y})}{h'(\bar{y})}E_i[\tilde{y}^*] + O^2\left(\{\tilde{p}_i, \epsilon_{i,k}\}_{i,\ k\neq i\in\{1,\cdots,N\}}\right),$$

and

$$\sum_{i=1}^N\left(\bar{p}_i\tilde{x}_i^*(s_i) + \tilde{p}_i\bar{x}_i\right) + \tilde{y}^* = O^2\left(\{\tilde{p}_i, \epsilon_{i,k}\}_{i,\ k\neq i\in\{1,\cdots,N\}}\right),$$

where $O^2(\cdot)$ denotes second or higher order terms. We then have[6]

$$\tilde{x}_i^{\text{Narrow}}(\tilde{p}_1,\cdots,\tilde{p}_N) = \sum_{k=1}^N\frac{\partial\tilde{x}_i^{\text{Narrow}}}{\partial\tilde{p}_k}\tilde{p}_k + O^2\left(\{\tilde{p}_k\}_{k\in\{1,\cdots,N\}}\right), \tag{47}$$

where the demand gradients $\left\{\frac{\partial\tilde{x}_i^{\text{Narrow}}}{\partial\tilde{p}_k}\right\}_{i,k\in\{1,\cdots,N\}}$ are given exactly by formulas in (44) and (45).

Consider the limit of part (iv) of Proposition 5 where $\omega_i \to 1$, when $\lambda_{j,i}\to 0$ for all $j\neq i$ and $\frac{\mu_y}{\kappa_y}\to 0$. In this limit, we have $\frac{\partial\tilde{x}_i^{\text{Narrow}}}{\partial\tilde{p}_i} \to \frac{\partial\tilde{x}_i^{\text{Explicit}}}{\partial\tilde{p}_i} = -\frac{\bar{x}_i}{\bar{p}_i}$ and $\frac{\partial\hat{x}_k^{\text{Narrow}}}{\partial\hat{p}_i}\to 0$ for $k\neq i$. In this limit, (47) becomes

$$\bar{p}_i\tilde{x}_i^{\text{Narrow}} + \tilde{p}_i\bar{x}_i = O^2\left(\{\tilde{p}_k\}_{k\in\{1,\cdots,N\}}\right).$$

[6]Note that $\frac{\partial\tilde{x}_i^{\text{Narrow}}}{\partial\tilde{p}_i}$ are the exact own-price demand gradient at the point of linearization/log-linearization.

Similar for the case of log-linearization in the main text: (14) and (15) become

$$-\kappa_i \hat{x}_i^* (s_i) = \hat{p}_i - E_i [\kappa_y \hat{y}^*] + O^2 \left(\{\hat{p}_i, \hat{\epsilon}_{i,k}\}_{i,\ k \neq i \in \{1, \cdots, N\}} \right),$$

and

$$\sum_{i=1}^{N} \mu_i (\hat{p}_i + \hat{x}_i^* (s_i)) + \mu_y \hat{y}^* = O^2 \left(\{\hat{p}_i, \hat{\epsilon}_{i,k}\}_{i,\ k \neq i \in \{1, \cdots, N\}} \right).$$

Similar to above, consider the limit in part (iv) of Proposition 5. We have

$$\hat{p}_i + \hat{x}_i^{\text{Narrow}} = O^2 \left(\{\hat{p}_k\}_{k \in \{1, \cdots, N\}} \right).$$

The Small Wage Elasticity of Daily Labor Supply.

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

Environment. To formalize, consider a decision maker whose utility is

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

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

Information. In this environment, each self $i \in \{1, \cdots, N\}$ should be interpreted as in charge of labor supply decisions for a day. Each self i of the narrow thinker perfectly knows the wage she faces w_i, and receives a noisy signal about each of the other selves' w_j. Specifically, for $i \in \{1, \cdots, N\}$, self i's information (signals) is given by $s_i = \left\{ \{s_{i,j}\}_{j \in \{1, \cdots, N\}} \right\}$, 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 $\sigma_{i,j}^2 > 0$. All ϵ are w are independent of each other. That is, for self $i \in \{1, \cdots, N\}$, her information is given by $s_i = \{w_i, s_{i,k}\}_{k \neq i,\ k \in \{1, \cdots, N\}}$.

Narrow thinker's behavior. Similar to the main text, I use a hat over a variable to denote its log-deviation from the point of log-linearization.[7] The optimal labor supply condition for each i and the

[7] 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.

budget constraint are given by

$$\kappa \hat{l}_i^* (s_i) = \hat{w}_i - \kappa_y E_i [\hat{y}^*], \tag{48}$$

$$\sum_{i=1}^{N} \mu_i \left(\hat{l}_i^* (s_i) + \hat{w}_i \right) = \hat{y}^*, \tag{49}$$

where $\kappa_y = -\frac{h''(\bar{y})\bar{y}}{h'(\bar{y})}$ and $\mu_i = \frac{\bar{w}_i \bar{l}_i}{\bar{y}}$ 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 (7), for each i, I define the narrow thinker's (log) labor supply function as $\hat{l}_i^{\text{Narrow}} (\hat{w}_1, \cdots, \hat{w}_N) \equiv E\left[\hat{l}_i^* (s_i) | \hat{w}_1, \cdots, \hat{w}_N \right]$. Compared to the standard frictionless case when each decision is made with perfect knowledge of all fundamentals (indexed by the superscript *Standard*, as above), one can then establish small wage elasticity of daily labor supply under narrow thinking.

Proposition 9 *For each i, the narrow thinker's labor supply l_i is smaller (larger) in response to positive (negative) shocks to w_i :*

$$\frac{\partial \hat{l}_i^{\text{Narrow}}}{\partial \hat{w}_i} = \omega_i \frac{\partial \hat{l}_i^{\text{Target}}}{\partial \hat{w}_i} + (1 - \omega_i) \frac{\partial \hat{l}_i^{\text{Standard}}}{\partial \hat{w}_i} \leq \frac{\partial \hat{l}_i^{\text{Standard}}}{\partial \hat{w}_i}, \tag{50}$$

where $s_i \in [0, 1]$ and l_i^{Target} captures a daily income target model. That is, there exists a daily income target m_i such that $w_i l_i = m_i$.

Proposition 9 expresses the narrow thinker's wage elasticity of daily labor supply as a weighted average between that in the standard labor supply theory (l_i^{Standard}) and that with a daily income target (l_i^{Target}). This leads to a smaller wage elasticity of daily labor supply under narrow thinking. To see the mechanism behind the small wage elasticity of daily labor supply, note that an increase in w_i increases l_i (a positive direct effect) and decreases l_j for $j \neq i$ (both in standard consumer theory and under narrow thinking). This is because the income effect of w_i on l_j (negative) and the substitution effect of w_i on l_j (negative) work in the same direction. The decrease of other l_j then further increases l_i (a positive indirect effect). Under narrow thinking, in response to an increase in w_i, the decision maker decreases labor supply on other days less. The indirect effect from this coordinated response is dampened, and the narrow thinker's l_i is smaller in response to the increase in w_i.

Economic implications. First, as Farber (2015) points out, daily income target model predicts negative wage elasticity of daily labor supply, which is inconsistent with the empirical evidence. By providing a smooth version of the daily income target model in (50), my approach can then explain the empirically documented positive, but small, wage elasticity of daily labor supply.

Second, in line with Proposition 3, the smaller wage elasticity of labor supply under narrow thinking is about response to temporary daily wage shocks. The narrow thinker's labor supply decision, average across days, as a function of the average wage can coincide with that in the standard benchmark. Such prediction is consistent with the larger wage elasticity of labor supply found in Fehr and Goette (2007) and Angrist, Caldwell and Hall (2021) based on wage variations at longer frequency.

Proof of Proposition 9. Similar to the proof of Proposition 10, from (48) and (49), we have, for all i and $k \neq i$,

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

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

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

$$\frac{\partial \hat{l}_i^{\text{Narrow}}}{\partial \hat{w}_i} = \frac{1}{\kappa} - \frac{\mu_i \frac{\kappa+1}{\kappa}}{\kappa \left(\frac{1}{\kappa_y} + \frac{\mu_i}{\kappa} + \sum_{j \neq i} \frac{\mu_j}{\kappa} \frac{\lambda_{j,i} \frac{1}{\kappa_y}}{\frac{1}{\kappa_y} + (1-\lambda_{j,i}) \frac{\mu_j}{\kappa}} \right)}. \tag{51}$$

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

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

where

$$\omega_i = 1 - \frac{\frac{1}{\kappa_y} + \sum_{j \neq i} \frac{\mu_j}{\kappa} \frac{\lambda_{j,i} \frac{1}{\kappa_y}}{\frac{1}{\kappa_y} + (1-\lambda_{j,i}) \frac{\mu_j}{\kappa}}}{\frac{1}{\kappa_y} + \frac{\mu_i}{\kappa} + \sum_{j \neq i} \frac{\mu_j}{\kappa} \frac{\lambda_{j,i} \frac{1}{\kappa_y}}{\frac{1}{\kappa_y} + (1-\lambda_{j,i}) \frac{\mu_j}{\kappa}}} \frac{\frac{1}{\kappa_y} + \frac{\mu_i}{\kappa} + \sum_{j \neq i} \frac{\mu_j}{\kappa}}{\frac{1}{\kappa_y} + \sum_{j \neq i} \frac{\mu_j}{\kappa}} \in [0, 1],$$

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

Appendix D: General Properties

Here, I provide more general results about the optimal behavior under narrow thinking: narrow thinking smoothly attenuates the interaction across decisions; when narrow thinking leads to over-reaction and when it leads to under-reaction.

Effective Attenuation of Interaction

From utility to the narrow thinker's optimal decision rules. Consider the general environment in Section 2. That is, the consumer's utility is given by is given by $u\left(\vec{x},\vec{\theta}\right)$ in (1), with $\vec{x}=(x_1,\cdots,x_N)$ and $\vec{\theta}=(\theta_1,\cdots,\theta_M)$.[8]

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

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

Lemma 5 *For each self $i\in\{1,\cdots,N\}$, her (log-linearized) optimal decision is given by*

$$\hat{x}_i^*(s_i)=\underbrace{E_i\left[\sum_{1\le k\le M}\psi_{i,k}\hat{\theta}_k\right]}_{\textit{direct effect}}+\underbrace{E_i\left[\sum_{j\neq i,1\le j\le N}\gamma_{i,j}\hat{x}_j^*(s_j)\right]}_{\textit{indirect effect}},\tag{52}$$

where $E_i[\cdot]=E[\cdot|s_i]$ denotes self i's belief, $\psi_{i,k}=-\frac{u_{x_i,\theta_k}}{u_{x_i,x_i}}\frac{\bar{\theta}_k}{\bar{x}_i}>0$, and $\gamma_{i,j}=-\frac{u_{x_i,x_j}}{u_{x_i,x_i}}\frac{\bar{x}_j}{\bar{x}_i}$.[9]

The first term in (52), which I call the direct effect, summarizes the fundamental $\vec{\theta}$'s direct influence on self i's decision, holding other selves' decisions fixed. For each $i\in\{1,\cdots,N\}$ and $k\in\{1,\cdots,M\}$, $\psi_{i,k}$ captures how the fundamental θ_k directly influences the optimal decision i.

The second term in (52), which I call the indirect effect, summarizes how other selves' decisions influence self i's decision. For each $i\neq j\in\{1,\cdots,N\}$, $\gamma_{i,j}$ summarizes how decision i is influenced by decision j. A positive (negative) $\gamma_{i,j}$ means that optimal decision i increases (decreases) with decision j. In fact, one can think of (52) as each self i's best response function in the equivalent game among multiple selves. It is akin to the best response function in a linear network game (Bergemann, Heumann and Morris, 2017; Golub and Morris, 2017), and $\Gamma=\{\gamma_{i,j}\}_{1\le i,j\le N}$ can be interpreted as the interaction matrix across different decisions.[10]

Effective attenuation of interaction. I now show that, in response to shocks to the fundamentals, narrow thinking smoothly attenuates the interaction across decisions. Similar to (7), for each

[8] As discussed above, constrained problems, such as consumer theory examples studied, can be transformed into the unconstrained problem here.

[9] Here, $u_{x_i,\theta_k}=\frac{\partial^2u\left(\bar{x}_1,\cdots,\bar{x}_1,\bar{\theta}_1,\cdots\bar{\theta}_M\right)}{\partial x_i\partial\theta_k}$, $u_{x_i,x_i}=\frac{\partial^2u\left(\bar{x}_1,\cdots,\bar{x}_1,\bar{\theta}_1,\cdots\bar{\theta}_M\right)}{(\partial x_i)^2}$, and $u_{x_i,x_j}=\frac{\partial^2u\left(\bar{x}_1,\cdots,\bar{x}_1,\bar{\theta}_1,\cdots\bar{\theta}_M\right)}{\partial x_i\partial x_j}$

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

$i \in \{1, \cdots, N\}$, I define the narrow thinker's (log-)decision function as

$$\hat{x}_i^{\text{Narrow}}\left(\hat{\theta}_1, \cdots, \hat{\theta}_M\right) \equiv E\left[\hat{x}_i^*\left(s_i\right) | \hat{\theta}_1, \cdots, \hat{\theta}_M\right], \tag{53}$$

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

Proposition 10 *In response to shocks to θ_k, the narrow thinker's decisions can be characterized by*

$$\begin{pmatrix} \frac{\partial \hat{x}_1^{\text{Narrow}}}{\partial \hat{\theta}_k} & \frac{\partial \hat{x}_2^{\text{Narrow}}}{\partial \hat{\theta}_k} & \cdots & \frac{\partial \hat{x}_N^{\text{Narrow}}}{\partial \hat{\theta}_k} \end{pmatrix}' = \left(\mathbb{I}_N - \tilde{\Gamma}_k\right)^{-1} \Psi_k, \tag{54}$$

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

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

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

The matrix $\tilde{\Gamma}_k$ captures the effective interaction across decisions in response to shocks to 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 smoothly attenuated by the factor $\lambda_{i,k} = \frac{\sigma_{\theta_k}^2}{\sigma_{\theta_k}^2 + \sigma_{i,k}^2} \in [0, 1]$, depending on the precision of the narrow thinker's information. That is, in response to shocks to θ_k, because self i has an imperfect perception of self j's decision, an one unit increase (decrease) in x_j only effectively increases (decreases) x_i by $\lambda_{i,k}\gamma_{i,j}$. It is as if self i cares less about the influence of other decisions, and she "thinks narrowly."[11]

The vector Ψ_k in (54) captures the direct effect of θ_k on each decision. As each self i may not perfectly know θ_k, The direct effect of θ_k on x_i can also be dampened. To isolate the attenuation of interaction across decisions (the friction of interest) from this dampening of direct effects, we can consider the case in which θ_k only directly influences x_k ($\psi_{i,k} = 0$ for $i \neq k$), and self k perfectly knows θ_k ($\lambda_{k,k} = 1$), e.g., the consumer theory example in Section 3. Then, the direct effects of θ_k on decisions are maintained,[12] and the sole friction comes from the effective attenuation of interaction above and the accompanying dampening of indirect effects.

It is true that a large number of parameters $\{\lambda_{i,k}\}_{\forall i,k}$ govern the effective attenuation of interaction. This reflects the complexity of each self's information environment, once we take into consideration bounded recall and selective retrieval from memory. In applications, one can use the following method to minimize

[11] 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.

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

the degree of freedom. One can let each self have perfect knowledge about one fundamental, i.e., p_i for each self i in the consumer theory example above, and set all other $\lambda_{i,k}$ to be the same λ. In fact, this is what I did in the calibration exercise in Corollary 2.

Over- and Under-reaction

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

Motivated by the above applications, I equate the number of fundamentals with the number of decisions ($M = N$), and impose that each self i perfectly knows fundamental θ_i ($\sigma_{i,i}^2 = 0$). In the consumer theory examples in Sections 3 and 4.1, this information structure just means that each self i perfectly knows the price of the good she buys, p_i. For notation simplicity, I also assume $\psi_{i,k} \geq 0$ for all i and k.[13]

Here, I study how each of the narrow thinker's decision x_i responds to shocks to θ_i, which she perfectly knows, e.g., the own-price demand elasticities in consumer theory. Below, I also study how each x_i responds to shocks to other θ_k.

From (52), we know the response of x_i to θ_i can be decomposed into a direct and an indirect effect.

$$\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{\theta}_i} = \underbrace{\psi_{i,i}}_{\text{Direct}} + \underbrace{\sum_{j \neq i} \gamma_{i,j} \frac{\partial E\left[E_i\left[\hat{x}_j\right] | \hat{\theta}_1, \cdots, \hat{\theta}_N\right]}{\partial \hat{\theta}_i}}_{\text{Indirect}}.$$

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

To formalize, we use $x_i^{\text{Standard,Ind}}$ to denote the indirect effect on decision i in (52), when each decision is made with perfect knowledge of all the fundamentals. As we normalize the direct effect of θ_i on x_i to be positive, the indirect effect of θ_i on x_i works in the same direction as the direct effect when $\frac{\partial \hat{x}_i^{\text{Standard,Ind}}}{\partial \hat{\theta}_i} > 0$. On the other hand, the indirect effect works in the opposite direction as the direct effect when $\frac{\partial \hat{x}_i^{\text{Standard,Ind}}}{\partial \hat{\theta}_i} < 0$.

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

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

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

[13] The analysis is the same when $\psi_{i,k} \leq 0$ for all i and k, e.g., in the context of consumer theory. One can just interpret $-\hat{p}_i$ as $\hat{\theta}_i$.

3) Substitutes with a single factor structure, i.e., there exists non-negative scalars $\{\rho_i, \Gamma_i, \Delta_i\}_{i=1}^N$ *such that* $\gamma_{i,j} = -\rho_i \Gamma_j$ *and* $\psi_{i,k} = \rho_i \Delta_k$ *for all* $j, k \neq i$ *and* $\rho_i \Gamma_i < 1$ *for all* i.[14'15]

Proposition 11 *Suppose Assumption 1 holds,*

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

$$\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{\theta}_i} \leq \frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{\theta}_i} \quad \forall i;$$

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

$$\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{\theta}_i} \geq \frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{\theta}_i} \quad \forall i.$$

The reason that additional conditions in Assumption 1 are required to establish Proposition 11 is because of the coexistence of opposing indirect effects. There could be some components of the indirect effect (e.g., through one decision x_{j_1}) that positively influence the optimal decision i and there could be some components of the indirect effect (e.g., through another decision x_{j_2}) that negatively influence the optimal decision i. Dampening of each component may not mean dampening of the net total. Each one of the additional conditions provided in Assumption 1 guarantees that one direction of the indirect effect dominates, and the *net* total of the indirect effect is dampened under narrow thinking. Depending on whether the *net* total of the indirect effect works in the same direction as the direct effect or not, narrow thinking then translates into over- or under- reaction.

In the proof of Proposition 11, I also show how all applications in Sections 3 and 4 satisfy Assumption 1. It is worth briefly explaining the third condition in Assumption 1, "Substitutes with a single factor structure." This condition is satisfied when the interaction across decisions comes from a common source, such as the budget constraint in the consumer theory example with income effects in Section 4.1.

Cross-elasticities

I now turn to how x_i responds to shocks to other fundamentals θ_k, for $i \neq k$.

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

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

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

[15]For case 3), note that the "single factor structure" only restricts $\psi_{i,k}$ for $i \neq k$. On the other hand, $\psi_{i,i}$, can be any non-negative scalar.

Similar to Proposition 11, the attenuation of cross-elasticity is not always true because of the coexistence of opposing effects. Dampening of each component may not mean dampening of the net total. The additional conditions guarantee that the *net* total of cross-elasticity is dampened.

Proof of Lemma 5. I start from the decision-specific optimality in (3), and take a first order condition:

$$E\left[\frac{\partial u}{\partial x_i}\left(x_1^*\left(s_1\right),\cdots,x_i,\cdots,x_N^*\left(s_N\right),\vec{\theta}\right)|s_i\right]=0\quad\forall i,s_i\in\Omega_i. \tag{55}$$

Log-linearize this condition, I arrive at (52).

Proof of Proposition 10. Based on Lemma 1, I use guess and verify approach to find the unique optimum. I conjecture the optimal decision rule for each self i, $\hat{x}_i^*\left(s_i\right)$, is linear in her signals,

$$\hat{x}_i^*\left(s_i\right)=\sum_{k=1}^{M}\alpha_{i,k}\hat{s}_{i,k}. \tag{56}$$

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

$$E_i\left[\hat{s}_{j,k}\right]=E_i\left[\hat{\theta}_k\right]=\lambda_{i,k}\hat{s}_{i,k},$$

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

$$E_i\left[\hat{x}_j^*\left(s_j\right)\right]=\sum_{k=1}^{N}\lambda_{i,k}\alpha_{j,k}\hat{s}_{i,k}. \tag{57}$$

Together with the optimal decision rule in (52) and the guess in (56), we have, for all i,

$$\hat{x}_i^*\left(s_i\right)=\sum_{k=1}^{M}\psi_{i,k}\lambda_{i,k}\hat{s}_{i,k}+\sum_{j\neq i}\gamma_{i,j}\sum_{k=1}^{M}\lambda_{i,k}\alpha_{j,k}\hat{s}_{i,k}.$$

For the guess in (56) to be valid, we then need to have, for all i,k,

$$\alpha_{i,k}=\psi_{i,k}\lambda_{i,k}+\sum_{j\neq i}\lambda_{i,k}\gamma_{i,j}\alpha_{j,k}. \tag{58}$$

(58) is satisfied when

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

This verifies that the guess in (56) indeed characterizes the narrow thinker's optimal decision rules. To prove Proposition 10, note that based on the definition in (53), for all i,

$$\hat{x}_i^{\text{Narrow}}\left(\vec{\theta}\right) = \sum_{k=1}^{N} \alpha_{i,k}\hat{\theta}_k.$$

Taking partial derivatives with respect to each θ_k then leads to Proposition 10.

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. From Lemma 5, we know (58) can then be re-written as

$$\lambda_{i,k}^{-1} u_{x_i,x_i}\bar{x}_i\alpha_{i,k} + \sum_{j\neq i} u_{x_i,x_j}\bar{x}_j\alpha_{j,k} = -u_{x_i,\theta_k}\bar{\theta}_k.$$

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).[16] Since U is negative definite,

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

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 11. I prove the Proposition 11 case by case. In the case (1) "Symmetry," there exists $\psi, \Psi > 0,\ \lambda \in (0,1)$, and $\gamma \in \left(-1, \frac{1}{N-1}\right)$, such that $\psi_{i,i} = \psi$, $\psi_{i,k} = \Psi$, $\gamma_{i,j} = \gamma$, and $\lambda_{i,j} = \lambda$ for all $j, k \neq i$.

From (54), we can express the own-sensitivity as

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

with $\frac{\partial \hat{x}_i^{\text{Standard,Ind}}}{\partial \hat{\theta}_i} = \gamma (N-1) \frac{\Psi + \gamma \psi}{1 - \gamma^2 (N-1) - \gamma (N-2)}$. Using the fact that $\lambda \in [0,1)$, $\psi, \Psi > 0$ and $\gamma \in \left(-1, \frac{1}{N-1}\right)$,[17] Proposition 11 follows directly.

In the case (2) "Complements," we have $\gamma_{i,j} \geq 0$ for all $i \neq j$ and $\sum_{j \neq i} \gamma_{i,j} < 1$ for all i. In this case,

[16] This uses $\bar{x}_i \neq 0$ for all i. Otherwise log-linearization will be invalid.

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

the game among multiple selves is solvable by iterating best response. From (54), we have

$$\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{\theta}_i} = \psi_{i,i} + \sum_{j\neq i} \gamma_{i,j}\psi_{j,i} + \sum_{j\neq i} \gamma_{i,j} \sum_{l\neq j} \lambda_{j,i}\gamma_{j,l}\psi_{l,i} + \cdots . \tag{59}$$

As each term in (59) is non-negative, the indirect effect always works in the same direction as the direct effect, and the result follows directly.

In the case (3) "Substitutes with a single factor structure," there exists non-negative scalars $\{\rho_i, \Gamma_i, \Delta_i\}_{i=1}^N$ such that $\gamma_{i,j} = -\rho_i\Gamma_j$ and $\psi_{i,k} = \rho_i \Delta_k$ for all $j, k \neq i$ and $\rho_i\Gamma_i < 1$ for all i. Define $\hat{y}_{-i}^{\text{Narrow}} = \sum_{j\neq i} \Gamma_j \hat{x}_j^{\text{Narrow}}$ for all i. Based on the proof of Proposition 10, we have, for all $i \neq k$,

$$\begin{aligned}
\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{\theta}_i} &= \psi_{i,i} - \rho_i \frac{\partial \hat{y}_{-i}^{\text{Narrow}}}{\partial \hat{\theta}_i} \\
\frac{\partial \hat{x}_j^{\text{Narrow}}}{\partial \hat{\theta}_i} &= \lambda_{j,i}\psi_{j,i} - \rho_j \lambda_{j,i} \left(\Gamma_i \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{\theta}_i} + \frac{\partial \hat{y}_{-i}^{\text{Narrow}}}{\partial \hat{\theta}_i} - \Gamma_j \frac{\partial \hat{x}_j^{\text{Narrow}}}{\partial \hat{\theta}_i} \right) \\
\frac{\partial \hat{y}_{-i}^{\text{Narrow}}}{\partial \hat{\theta}_i} &= \sum_{j\neq i} \Gamma_j \frac{\partial \hat{x}_j^{\text{Narrow}}}{\partial \hat{\theta}_i}.
\end{aligned}$$

Together, we have

$$\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{\theta}_i} = \psi_{i,i} - \rho_i \left(\Delta_i - \Gamma_i \psi_i\right) \frac{\sum_{j\neq i} \frac{\lambda_{j,i}\rho_j\Gamma_j}{1-\lambda_{j,i}\rho_j\Gamma_j}}{1 + (1-\rho_i\Gamma_i) \sum_{j\neq i} \frac{\lambda_{j,i}\rho_j\Gamma_j}{1-\rho_j\lambda_{j,i}\Gamma_j}},$$

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

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

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

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

ii) For the environment in Section 4.1, based on each self's optimal consumption rule in (19), it falls into case 3) of Assumption 1, with $\rho_i = \frac{\frac{\kappa_y}{\kappa_i}}{1+\frac{\kappa_y}{\mu_y}\frac{\mu_i}{\kappa_i}}$ and $\Gamma_i = \Delta_i = \frac{\mu_i}{\mu_y}$.[18]

iii) For the environment in Proposition 6, from (32) and (33), we know each self's optimal consumption rule can be written as

$$\hat{x}_i^*(s_i) = \frac{\frac{1}{\kappa_i}}{1 + \frac{\kappa_y}{\kappa_i}\frac{\mu_i}{\mu_y}} \hat{\varphi}_i - E_i \left[\sum_{j\neq i} \frac{\frac{\kappa_y}{\kappa_i}\frac{\mu_j}{\mu_y}}{1 + \frac{\kappa_y}{\kappa_i}\frac{\mu_i}{\mu_y}} \hat{x}_j^*(s_j) \right].$$

[18]Here I interpret $-\hat{p}_i$ as $\hat{\theta}_i$.

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

iv) For the environment in Proposition 7, from (34) and (35), we know each self's optimal consumption rule can be written as

$$\hat{x}_i^*(s_i) = E_i\left[\sum_{j=1}^{N}\frac{\frac{\kappa_y}{\mu_y}\frac{\mu_j^w}{\kappa_i}}{1+\frac{\kappa_y}{\mu_y}\frac{\mu_i}{\kappa_i}}\hat{w}_j - \sum_{j\neq i}\frac{\frac{\kappa_y}{\mu_y}\frac{\mu_j}{\kappa_i}}{1+\frac{\kappa_y}{\mu_y}\frac{\mu_i}{\kappa_i}}\hat{x}_j^*(s_j)\right],$$

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

It is worth noting that for case 3) of Assumption 1, the "single factor structure" only restricts $\psi_{i,k}$ for $i \neq k$. On the other hand, $\psi_{i,i}$, can be any non-negative scalar.

Proof of Proposition 12. We prove the Proposition 12 case by case. In the case (1) "Symmetry," there exists $\psi, \Psi > 0,\ \lambda \in (0,1)$, and $\gamma \in \left(-1, \frac{1}{N-1}\right)$, such that $\psi_{i,i} = \psi$, $\psi_{i,k} = \Psi$, $\gamma_{i,j} = \gamma$, and $\lambda_{i,j} = \lambda$ for all $j, k \neq i$.

From (54), we can express the cross-sensitivity as

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

Using the fact that $\lambda \in [0,1)$, $\psi, \Psi > 0$ and $\gamma \in \left(-1, \frac{1}{N-1}\right)$,[19] Proposition 12 follows directly.

In the case (2) "Complements," we have $\gamma_{i,j} \geq 0$ for all $i \neq j$ and $\sum_{j\neq i}\gamma_{i,j} < 1$ for all i. In this case, the game among multiple selves is solvable by iterating best response. From (54), we have

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

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

Appendix E: Additional Applications

Myopic Loss Aversion.

Another behavior often connected to narrow bracketing is the decision maker's aversion to combining small and favorable gambles (Samuelson, 1963). The existing explanations of this behavior, such as Benartzi and Thaler (1995), Barberis, Huang and Thaler (2006), and Rabin and Weizsacker (2009), contains two elements: first, the decision maker suffers from loss aversion; second, she decides on each gamble in isolation.[20] Narrow

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

[20]Loss aversion alone is not enough, as the decision maker can combine independent and favorable gambles to avoid the loss.

thinking provides a formal explanation of such myopic loss aversion, without directly requiring the decision maker to decide on each gamble in isolation.

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

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

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

For each gamble $i \in \{1, 2\}$, there is a 50% chance that it turns out to be a loss of $r_i = -1$ dollars, and a 50% chance that it turns out to be a gain of $r_i = 1 + \mu_i$ dollars, where $\mu_i = \mu\epsilon_i > 0$ is a random variable, with $\mu \sim \log\mathcal{N}\left(\log(\mu^{avg}) - \frac{1}{2}\sigma^2, \sigma^2\right)$ and $\epsilon_i \sim \log\mathcal{N}\left(-\frac{1}{2}\sigma_\epsilon^2, \sigma_\epsilon^2\right)$ for $i \in \{1, 2\}$. μ, ϵ_1, and ϵ_2 are independent of each other. We then have $E[e^{\mu_i}] = \mu^{avg} > 0$ for each $i \in \{1, 2\}$. Moreover, whether each gamble turns out to be a gain or a loss is independent of that of the other gamble and the returns of gambles.

First consider a frictionless decision maker (indexed by "Standard") who makes each gambling decisions with knowledge of both μ_1 and μ_2. She can coordinate her gambling decisions by combining two gambles: if one gamble turns out to be a loss and another gamble turns out to be a gain, she will not suffer from loss aversion. In fact, she either invests in both gambles or does not invest in any of them:

$$\begin{cases} x_1^{\text{Standard}} = x_2^{\text{Standard}} = 1 & \text{if } \mu(\epsilon_1 + \epsilon_2) > \delta - 1 \\ x_1^{\text{Standard}} = x_2^{\text{Standard}} = 0 & \text{if } \mu(\epsilon_1 + \epsilon_2) < \delta - 1 \end{cases}.$$

Now let us turn to the narrow thinker. Each self $i \in \{1, 2\}$ is in charge of deciding whether to bet on gamble i or not, i.e. the decision on $x_i \in \{0, 1\}$. Each self i perfectly knows μ_i, but does not know μ_{-i}. She instead needs to use μ_i to infer μ_{-i}.[21] As a result, the narrow thinker cannot perfectly coordinate their gambling decisions. This difficulty in coordinating decisions makes it harder for the narrow thinker to enjoy the benefits of combining two gambles together. This leads to a lower probability of investing in each gamble and provides a model of myopic loss aversion.

In particular, Figure 2 plots the probability of investing in each gamble, i.e., $\text{Prob}(x_i = 1)$, for the standard decision maker and the narrow thinker. It plots the probability of gambling as a function of the noise in individual return σ_ϵ, while setting $\delta = 2.25$ (Tversky and Kahneman, 1992), $\mu^{avg} = 1$, and normalizing $\sigma = 1$. We can see that the narrow thinker has a lower probability of investing in each gamble. When σ_ϵ is larger, the information sets of two selves' of the narrow thinker are more distant, and the narrow thinker's probability of investing deviates further from the standard case.

[21] In the current set up, each self's only information about μ_{-i} comes from her own gamble μ_i. She does not receives an additional signal about μ_{-i}. This helps facilitate the characterization of the narrow thinker's behavior below: each self's gambling strategy can be characterized by a single-dimension cutoff strategy.

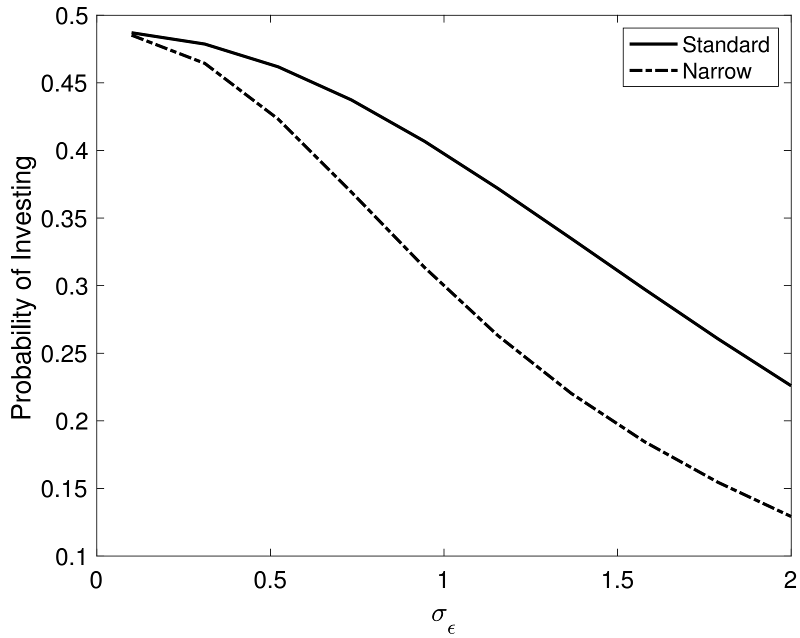

Figure 2: The lower bound of the degree of mental accounting.

Characterizing the narrow thinker's behavior. I use guess and verify approach.[22] I guess the threshold equilibrium for the narrow thinker takes the following form: for $i \in \{1, 2\}$,

$$\begin{cases} x_i^{\text{Narrow}} = 1 & \text{if } \mu_i > \bar{\mu} \\ x_i^{\text{Narrow}} = 0 & \text{if } \mu_i < \bar{\mu} \end{cases}.$$

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

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

If she invests, her utility is given by

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

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

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

Based on the distribution of μ_i and μ_{-i}, I can find the unique solution for the threshold μ^* in the above

[22]In fact, such an equilibrium is the unique rationalization outcome of the game between the two selves.

expression. I can then calculate the probability of investing in each gamble for the narrow thinker in Figure 2.

Testable prediction of the narrow thinking approach. Consistent with the discussion in the main text, a testable prediction of the narrow thinking approach to myopic loss aversion is: the decision maker will be more reluctant to gamble if she makes different gambling decisions separately, based on different states of mind. This prediction is consistent with the finding in Redelmeier and Tversky (1992). They find that, when participants make several gambling decisions sequentially, they are less willing to gamble compared to the case when they make these gambling decisions together.

Neglect of "Adding-up" Effects.

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

Specifically, the decision maker's utility is given by

$$\sum_{i=1}^{N} \varphi_i v\left(x_i\right) - c\left(\sum_{i=1}^{N} x_i\right), \tag{60}$$

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

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

Similar to the main analysis, I define the narrow thinker's (log) decision as a function of the fundamentals as $\hat{x}_i^{\text{Narrow}}\left(\hat{\varphi}_1, \cdots, \hat{\varphi}_N\right) \equiv E\left[\hat{x}_i^*\left(s_i\right) | \hat{\varphi}_1, \cdots, \hat{\varphi}_N\right]$. 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.

Proposition 13 *For each i, the narrow thinker increases (decreases) her smoking on occasion 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, she has imperfect perception about how other x_js will respond to the common shock. 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 cost function is convex. As a result, the direct effect and the indirect effect of the common shock work in opposite directions, and narrow thinking leads to over-reaction. Intuitively, when smoking becomes more attractive, the narrow thinker under-estimates how other selves will increase smoking, neglects the adding-up costs, and smokes excessively.

More generally, in response to common shocks to the fundamental, if different selves' decisions are strategic substitutes, narrow thinking leads to overreaction relative to the frictionless benchmark. This case arises when the decision maker faces convex add-up costs (e.g., smoking) or concave add-up benefits (e.g., demand for variety in choices). On the other hand, if different selves' decisions are strategic complements, narrow thinking leads to under reaction relative to the frictionless benchmark. This case arises when the decision maker faces convex add-up benefits (e.g., skill acquisition) and concave add-up costs (e.g., habituation). See the proof of Proposition 13 for details.[23]

Proof of Proposition 13. Given the environment, the optimal decision rule for each i is

$$\hat{x}_i^*(s_i) = E_i\left[\psi\hat{\varphi}_i - \gamma\sum_{j\neq i}\hat{x}_j^*(s_j)\right], \tag{61}$$

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

$$E_i[\hat{\varphi}_i] = \hat{\varphi}_i \quad \forall i, \tag{62}$$

[23] This relationship between strategic complementarity/substitutability and under-/over-reaction under narrow thinking only holds in response to a common shock. If the shock is idiosyncratic, as the case in Sections 3 - 4, we should rely on Proposition 11 to predict whether narrow thinking leads to over-reaction or under-reaction in a given environment.

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

$$\begin{aligned}E_i\left[\hat{\varphi}_j\right]&=E_i\left[\hat{\varphi}\right]+\frac{\sigma_\epsilon^{-2}}{\sigma_\delta^{-2}+\sigma_\epsilon^{-2}}\left(\hat{s}_{i,j}-E_i\left[\hat{\varphi}\right]\right)\\&\equiv\lambda\hat{s}_{i,j}+\mu\hat{\varphi}_i+\omega\sum_{l\neq i,j}\hat{s}_{i,l}\quad\forall i\neq j,\end{aligned}\tag{63}$$

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

Similar to the proof of Proposition 10, as the optimal decision rule (61) is linear and all variables are distributed normally, for all i, $\hat{x}_i^*(s_i)$ is linear in its signal and $\hat{x}_i^{\text{Narrow}}(\hat{\varphi}_1,\cdots,\hat{\varphi}_N)$ is linear in all $\hat{\varphi}$s. From condition (61) and the fact that the noise in each self's private signal is not predictable, we have

$$\hat{x}_i^*(s_i)=\psi\hat{\varphi}_i-\gamma\sum_{j\neq i}\hat{x}_j^{\text{Narrow}}\left(E_i\left[\hat{\varphi}_1\right],\cdots E_i\left[\hat{\varphi}_N\right]\right),$$

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 (62) and (63), averaging across noises in the realizations of signals, and taking partial derivatives with respect to each $\hat{\varphi}_j$, we have

$$\begin{aligned}\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_i}&=\psi-\gamma\sum_{j\neq i}\frac{\partial\hat{x}_j^{\text{Narrow}}}{\partial\hat{\varphi}_i}-\mu\gamma\sum_{j\neq i}\sum_{l\neq i}\frac{\partial\hat{x}_j^{\text{Narrow}}}{\partial\hat{\varphi}_l}\quad\forall i,\\\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_k}&=-\lambda\gamma\sum_{j\neq i}\frac{\partial\hat{x}_j^{\text{Narrow}}}{\partial\hat{\varphi}_k}-\omega\gamma\sum_{j\neq i}\sum_{l\neq k,i}\frac{\partial\hat{x}_j^{\text{Narrow}}}{\partial\hat{\varphi}_l}\quad\forall i,k.\end{aligned}$$

Using symmetry, we know $\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_i}$ are equal for each i and $\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_k}$ are equal for each $i\neq k$, we then have,

$$\begin{aligned}\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_i}&=\psi-\gamma(N-1)\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_k}-\mu\gamma\left((N-1)\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_i}+(N-1)(N-2)\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_k}\right),\\\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_k}&=-\lambda\gamma\left\{(N-2)\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_k}+\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_i}\right\}-\omega\gamma\left((N-2)\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_i}+(N-2)^2\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_k}\right).\end{aligned}$$

Collecting terms, we have

$$\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_i}=\frac{\psi}{1+\mu\gamma(N-1)-\frac{\gamma^2(N-1)(1+\mu(N-2))(\lambda+\omega(N-2))}{1+\lambda\gamma(N-2)+\omega\gamma(N-2)^2}}$$

$$\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_k}=-\frac{\gamma(\lambda+\omega(N-2))}{1+\gamma(N-2)(\lambda+\omega(N-2))}\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_i}.$$

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

$$\begin{aligned}\frac{d\hat{x}_i^{\text{Narrow}}}{d\hat{\varphi}} &= \frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_i} + (N-1)\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_k} = \frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{\varphi}_i}\left(1 - \frac{\gamma\left(\lambda+\omega\left(N-2\right)\right)\left(N-1\right)}{\left(1+\gamma\left(N-2\right)\left(\lambda+\omega\left(N-2\right)\right)\right)}\right) \\ &= \frac{\psi\left(1-\gamma\left(\lambda+\omega\left(N-2\right)\right)\right)}{\left(1+\mu\gamma\left(N-1\right)\right)\left(1+\gamma\left(N-2\right)\left(\lambda+\omega\left(N-2\right)\right)\right)-\gamma^2\left(N-1\right)\left(1+\mu\left(N-2\right)\right)\left(\lambda+\omega\left(N-2\right)\right)} \\ &= \frac{\psi\left(1-\gamma\left(\lambda+\omega\left(N-2\right)\right)\right)}{1+\mu\gamma\left(N-1\right)+\gamma\left(N-2-\gamma\left(N-1\right)\right)\left(\lambda+\omega\left(N-2\right)\right)}.\end{aligned}$$

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

$$\begin{aligned}\frac{d\hat{x}_i^{\text{Narrow}}}{d\hat{\varphi}} &> \frac{\psi\left(1-\gamma t\right)}{1+\gamma\left(N-1\right)\left(1-t\right)+\gamma\left(N-2-\gamma\left(N-1\right)\right)t} \\ &= \frac{\psi}{1+\gamma\left(N-1\right)} = \frac{d\hat{x}_i^{\text{Standard}}}{d\hat{\varphi}}.\end{aligned}$$

Comment. In the proof, what drives the over-reaction is that different decisions are strategic substitutes (from (61), this means $\gamma>0$). If different decisions are strategic complements ($\gamma<0$), we have under-reaction.

$$\frac{d\hat{x}_i^{\text{Narrow}}}{d\hat{\varphi}} < \frac{\psi\left(1-\gamma t\right)}{1+\gamma\left(N-1\right)\left(1-t\right)+\gamma\left(N-2-\gamma\left(N-1\right)\right)t} = \frac{\psi}{1+\gamma\left(N-1\right)} = \frac{d\hat{x}_i^{\text{Standard}}}{d\hat{\varphi}}.$$

Appendix F: Endogenous Narrow Thinking

Revisiting the Simple Consumer Theory Example

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

$$u\left(x_1,x_2,p_1,p_2\right)=v\left(x_1,x_2\right)+w-p_1x_1-p_2x_2. \tag{64}$$

Here, for an analytical solution, I assume v is strictly concave and quadratic. For a more general utility function, e.g., in Section 3, one can work with a log-quadratic approximation and the analysis will be the same.

At the information side, I do not directly impose that each self has perfect knowledge of p_i. Instead, I let the decision maker choose endogenously the precision of each self's signal about p_1 and p_2. Specifically, each potential signal $\boldsymbol{s_i}=\{\boldsymbol{s_{i,1}},\boldsymbol{s_{i,2}}\}\in\boldsymbol{\Omega_i}$ for decision i consists of a noisy signal about $\boldsymbol{p_1}$, $\boldsymbol{s_{i,1}}=\boldsymbol{p_1}+\boldsymbol{\epsilon_{i,1}}$, and a noisy signal about $\boldsymbol{p_2}$, $\boldsymbol{s_{i,2}}=\boldsymbol{p_2}+\boldsymbol{\epsilon_{i,2}}$. All $\boldsymbol{\epsilon}$ and $\boldsymbol{p}$ are normally distributed and independent of fundamentals and each other. The variances of the noises in these signals are free to choose, subject to the cognitive cost in (25).

In the optimum, let $\left(\sigma_{i,k}^*\right)^2$ denote the variance of the noise of self i's signal about $\boldsymbol{p_k}$ and $\lambda_{i,k}^*=$

$\frac{\sigma_{p_k}^2}{\sigma_{p_k}^2+\left(\sigma_{i,k}^*\right)^2}$ denote the precision of self i's signal about p_k. We have:

Proposition 14 *In the optimum in (25):*

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

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

Proof of Proposition 14. As discussed in the main text, the problem in (25) can be divided into two subproblems, the optimal information choice subject to the cognitive cost, and the optimal decisions *given* the chosen information. From the utility (64), given any chosen information $\{s_i\}_{i=1}^2$, the optimal decision rule $\{x_i^*(s_i)\}_{i=1}^2$ for $i \in \{1,2\}$ can be characterized by

$$E_i\left[\frac{\partial v}{\partial x_i}\left(x_i^*(s_i), x_{-i}^*(s_{-i})\right) - p_i\right] = 0. \tag{65}$$

Using law of iterated expectations, we henceforth have

$$E\left[x_i^*(s_i)\frac{\partial v}{\partial x_i}\left(x_i^*(s_i), x_{-i}^*(s_{-i})\right) - p_i x_i^*(s_i)\right] = 0.$$

Substituting into the decision maker's utility function, and using the fact that v is quadratic, the optimal information choice in (25) is then equivalent to[24]

$$\max_{\{\boldsymbol{s_i}\in\boldsymbol{\Omega_i}\}_{i=1}^2} -\frac{1}{2}E\left[p_1 x_1^*(s_1) + p_2 x_2^*(s_2)\right] - \phi\sum_{i=1}^N I\left(\boldsymbol{s_i};\vec{\boldsymbol{\theta}}\right) \tag{66}$$
$$s.t. \quad x_i^*(s_i) \text{ satisfy (65).}$$

Now, given the $\boldsymbol{\Omega_i}$ specified above, any $\boldsymbol{s_i} = \{\boldsymbol{s_{i,1}}, \boldsymbol{s_{i,2}}\}$ takes the form of $\boldsymbol{s_{i,1}} = \boldsymbol{p_1} + \boldsymbol{\epsilon_{i,1}}$ and $\boldsymbol{s_{i,2}} = \boldsymbol{p_2} + \boldsymbol{\epsilon_{i,2}}$, 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 $\boldsymbol{p}$s are independent of each other. Similar to the proof of Proposition 10, we have

$$\begin{pmatrix} \frac{\partial E[x_1^*(s_1)|p_1,p_2]}{\partial p_1} \\ \frac{\partial E[x_2^*(s_2)|p_1,p_2]}{\partial p_1} \end{pmatrix} = \left(\mathbb{I}_N - \begin{pmatrix} 1 & \lambda_{1,1} \\ \lambda_{2,1} & 1 \end{pmatrix} \circ \Gamma\right)^{-1} \begin{pmatrix} -\lambda_{1,1}\psi_1 \\ 0 \end{pmatrix},$$

[24] In (66), I omit constants that are independent of the information choice.

and

$$\begin{pmatrix} \frac{\partial E[x_1^*(s_1)|p_1,p_2]}{\partial p_2} \\ \frac{\partial E[x_2^*(s_2)|p_1,p_2]}{\partial p_2} \end{pmatrix} = \left(\mathbb{I}_N - \begin{pmatrix} 1 & \lambda_{1,2} \\ \lambda_{2,2} & 1 \end{pmatrix} \circ \Gamma \right)^{-1} \begin{pmatrix} 0 \\ -\lambda_{2,2}\psi_2 \end{pmatrix},$$

where $\lambda_{i,j} = \frac{\sigma_{p_j}^2}{\sigma_{p_j}^2+\sigma_{i,j}^2} \in (0,1]$, $\psi_i = -\left(\frac{\partial^2 v}{\partial x_i^2}\right)^{-1} > 0$, $\Gamma = \{\gamma_{i,j}\}_{1\le i,j\le N}$, and $\gamma_{i,j} = -\frac{\partial^2 v}{\partial x_i \partial x_j}/\frac{\partial^2 v}{\partial x_i^2}$. The problem in (66) then becomes

$$\max_{0\le\lambda_{i,j}\le1} g\left(\{\lambda_{i,j}\}\right) - \phi h\left(\{\lambda_{i,j}\}\right), \tag{67}$$

where $g\left(\{\lambda_{i,j}\}\right) \equiv \frac{1}{2}\psi_1 \frac{\lambda_{1,1}}{1-\lambda_{1,1}\lambda_{2,1}\gamma_{1,2}\gamma_{2,1}}\sigma_{p_1}^2 + \frac{1}{2}\psi_2 \frac{\lambda_{2,2}}{1-\lambda_{2,2}\lambda_{1,2}\gamma_{1,2}\gamma_{2,1}}\sigma_{p_2}^2$ and $h\left(\{\lambda_{i,j}\}\right) \equiv \sum_{1\le i,j\le 2}\frac{1}{2}log_2\left(\frac{1}{1-\lambda_{i,j}}\right)$ and I use the fact that all $\boldsymbol{\epsilon}$s and $\boldsymbol{p}$s are independent of each other and $I\left(\boldsymbol{s_i};\vec{\boldsymbol{p}}\right) = \frac{1}{2}log_2\left(\frac{1}{1-\lambda_{i,1}}\right) + \frac{1}{2}log_2\left(\frac{1}{1-\lambda_{i,2}}\right)$.

Now we prove Proposition 14. If, in the optimum, $\lambda_{1,1}^* \le \lambda_{2,1}^*$, $\frac{\partial g\left(\{\lambda_{i,j}^*\}_{1\le i,j\le 2}\right)}{\partial\lambda_{1,1}^*} > \frac{\partial g\left(\{\lambda_{i,j}^*\}_{1\le i,j\le 2}\right)}{\partial\lambda_{2,1}^*}$ 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 (67):

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

Therefore, $\lambda_{1,1}^* > \lambda_{2,1}^*$. Similarly, we can prove $\lambda_{2,2}^* > \lambda_{1,2}^*$.

Endogenizing the Degree of Narrow Bracketing

As discussed after Proposition 2, the endogenous narrow thinking problem considered here also generates predictions on when the degree of narrow bracketing is larger. If the price i is less volatile (a lower $\sigma_{p_i}^2$), the other self will pay less attention to p_i (a lower $\lambda_{-i,i}$), as the utility loss of narrow thinking is smaller. The decision maker will then narrowly bracket more in response to p_i (a higher ω_i).

Specifically, consider the simple consumer theory example above in Proposition 14. Similar to Proposition 2, I set $\sigma_{i,i}^* = 0$ and $\lambda_{i,i}^* = 1$. That is, each self $i \in \{1,2\}$ effortlessly knows the price of the good she buys. The "endogeneity" in this problem is that each self $-i$ chooses how much she knows about price p_i. I can then establish:

Proposition 15 *If $\gamma_{1,2}\gamma_{2,1} < \frac{1}{2}$ and ϕ is large enough,*[25] *for $i \in \{1,2\}$, $\lambda_{-i,i}^*$ increases with $\sigma_{p_i}^2$ and decreases with ϕ.*

Proof of Proposition 15. From (67), the problem here becomes

[25] Those conditions guarantee the concavity of the endogenous narrow thinking problem.

$$\max_{0\leq\lambda_{1,2},\lambda_{2,1}\leq 1}\frac{1}{2}\psi_1\frac{1}{1-\lambda_{2,1}\gamma_{1,2}\gamma_{2,1}}\sigma_{p_1}^2+\frac{1}{2}\psi_2\frac{1}{1-\lambda_{1,2}\gamma_{1,2}\gamma_{2,1}}\sigma_{p_2}^2-\frac{\phi}{2}\left[log_2\left(\frac{1}{1-\lambda_{1,2}}\right)+log_2\left(\frac{1}{1-\lambda_{2,1}}\right)\right].$$

Using $f(\lambda_{1,2},\lambda_{2,1})$ to denote the above objective, we have

$$\frac{\partial f(\lambda_{1,2},\lambda_{2,1})}{\partial\lambda_{1,2}}=\frac{1}{2}\psi_2\sigma_{p_2}^2\frac{\gamma_{1,2}\gamma_{2,1}}{(1-\lambda_{1,2}\gamma_{1,2}\gamma_{2,1})^2}-\frac{\phi}{2}\frac{1}{\ln 2\,(1-\lambda_{1,2})}.$$

When ϕ is large enough, $\frac{\partial f(\lambda_{1,2},\lambda_{2,1})}{\partial\lambda_{1,2}}$ is decreasing in $\lambda_{1,2}\in[0,1)$ so f is concave in $\lambda_{1,2}\in[0,1)$. We can then use FOC to characterize the optimal $\lambda_{1,2}$:[26]

$$\frac{1}{2}\psi_2\sigma_{p_2}^2\frac{\gamma_{1,2}\gamma_{2,1}}{\left(1-\lambda_{1,2}^*\gamma_{1,2}\gamma_{2,1}\right)^2}=\frac{\phi}{2}\frac{1}{\ln 2\left(1-\lambda_{1,2}^*\right)}.$$

Let $g\left(\gamma_{1,2}\gamma_{2,1},\lambda_{1,2}^*\right)\equiv\frac{\left(1-\gamma_{1,2}\gamma_{2,1}\lambda_{1,2}^*\right)^2}{1-\lambda_{1,2}^*}$. We have $\frac{\partial g\left(\gamma_{1,2}\gamma_{2,1},\lambda_{1,2}^*\right)}{\partial\lambda_{1,2}^*}=\frac{\left(1-\gamma_{1,2}\gamma_{2,1}\lambda_{1,2}^*\right)}{1-\lambda_{1,2}^*}\left[\frac{1+\gamma_{1,2}\gamma_{2,1}\lambda_{1,2}^*-2\gamma_{1,2}\gamma_{2,1}}{1-\lambda_{1,2}^*}\right]>0$. As a result, $\lambda_{1,2}^*$ increases with $\sigma_{p_2}^2$. Similarly, we can prove that $\lambda_{2,1}^*$ increases with $\sigma_{p_1}^2$. The comparative statics with respect to ϕ also follow directly.

Endogenizing the Degree of Mental Accounting

As discussed after Proposition 5, the endogenous narrow thinking problem considered here also generates predictions on when the degree of mental accounting (ω_i in Propositions 4 and 5) is larger.

To illustrate, consider the $N=2$ version of (13), the utility used in Section 4.1. Consistent with log-linearization used in the main analysis, here I consider a log-quadratic approximation of the utility function and the budget constraint. One can then analyze the endogenous narrow thinking problem in (25) analytically.

At the information side, same as Section 4.1, each self $i\in\{1,2\}$ perfectly knows $\boldsymbol{p_i}$ but receives a noisy signal about each of the other $\boldsymbol{p_{-i}}$: $\boldsymbol{s_{i,-i}}=\boldsymbol{p_{-i}\epsilon_{i,-i}}$, with $\boldsymbol{p_i}\sim\log\mathcal{N}\left(\log\bar{p}_i,\sigma_{p_i}^2\right)$ and $\boldsymbol{\epsilon_{i,-i}}\sim\log\mathcal{N}\left(0,\sigma_{i,-i}^2\right)$. All ϵs and ps are independent of each other. But now, the variances of the noises $\left\{\sigma_{i,-i}^2\right\}_{i\in\{1,2\}}$ are free to choose, subject to the cognitive cost in (25). Similar to Proposition 14, I let $\left(\sigma_{i,-i}^*\right)^2$ denote the variance of the noise of self i's signal about $\boldsymbol{p_{-i}}$ and $\lambda_{i,-i}^*=\frac{\sigma_{p_{-i}}^2}{\sigma_{p_{-i}}^2+\left(\sigma_{i,-i}^*\right)^2}$ denote the precision of self i's signal about $\boldsymbol{p_{-i}}$ in the optimum.

I can then establish:

Proposition 16 *Suppose that ϕ is large enough and $\left(\frac{\mu_y}{\kappa_y}\right)\left(\frac{\mu_1}{\kappa_1}+\frac{\mu_2}{\kappa_2}+\frac{\mu_y}{\kappa_y}\right)>\frac{\mu_1}{\kappa_1}\frac{\mu_2}{\kappa_2}$ for all $i\in\{1,2\}$.*[27]

[26] It is possible that, when ϕ is very large, an interior solution does not exist and $\lambda_{1,2}^*=\lambda_{2,1}^*=0$. Even considering this case, Proposition 15 still holds with "weakly" increasing.

[27] Those conditions guarantee the concavity of the endogenous narrow thinking problem.

Self $-i$ will choose to know less about p_i (a smaller $\lambda^*_{-i,i}$) and the degree of mental accounting (ω_i) will becomes larger when

(i) price p_i is less volatile ($\sigma^2_{p_i}$ is lower);

(ii) the cognitive cost ϕ is larger;

(iii) $\bar{\eta}w$ is smaller, where $\bar{\eta} = h'(\bar{y}) = \frac{\partial v(\bar{x}_1, \bar{x}_2)}{\partial x_i} / \bar{p}_i$ captures the marginal value of money (at the prior) and w is the decision maker's wealth.

Part (i) of Proposition 16 means: if the price of good i is less volatile (a lower $\sigma^2_{p_i}$), the other selves will choose to know less about p_i (a lower $\lambda_{-i,i}$), as the utility loss of narrow thinking is smaller. Part (ii) of Proposition 16 means: a larger cognitive cost ϕ will also lead to a lower $\lambda_{-i,i}$ and more mental accounting.

Part (iii) of Proposition 16 answers the question whether a richer decision maker will have a lower $\lambda_{-i,i}$ and exhibit more mental accounting. It turns out that there are countervailing channels: a richer decision maker has a smaller marginal utility of money $\bar{\eta}$, which pushes towards a lower $\lambda_{-i,i}$; but a richer decision maker also has a larger wealth w and buy more goods, which pushes towards a higher $\lambda_{-i,i}$. Which channel dominates depends on whether $\bar{\eta}w$ increases with wealth or not.

Proof of Proposition 16. Similar to the main text, I use a hat over a variable to denote its log-deviation from the point where each price is fixed at $\bar{p}_i$ and each decision is made with perfect knowledge of all prices.

Here, I consider a log-quadratic approximation of the utility function and the budget constraint. The utility in (13) becomes[28]

$$\sum_{i=1}^{2} \frac{\bar{x}_i^{1-\kappa_i}}{1-\kappa_i} \left[(1-\kappa_i)\hat{x}_i + \frac{1}{2}(1-\kappa_i)^2 \hat{x}_i^2 \right] + \bar{y}h'(\bar{y})\hat{y} + \frac{1}{2}\left[h''(\bar{y})(\bar{y})^2 + h'(\bar{y})\bar{y} \right] \hat{y}^2,$$

while the budget constraint becomes

$$\sum_{i=1}^{2} \bar{p}_i \bar{x}_i \left[\hat{p}_i + \hat{x}_i + \frac{1}{2}(\hat{p}_i + \hat{x}_i)^2 \right] + \bar{y}\left[\hat{y} + \frac{1}{2}(\hat{y})^2 \right] = 0.$$

Together and dropping terms with orders higher than 2, the decision maker is essentially maximizing the following unconstrained objective, which is nested by the problem (25):

$$\sum_{i=1}^{2} \frac{\bar{x}_i^{1-\kappa_i}}{1-\kappa_i} \left[(1-\kappa_i)\hat{x}_i + \frac{1}{2}(1-\kappa_i)^2 \hat{x}_i^2 \right] - h'(\bar{y}) \left[\sum_{i=1}^{2} \bar{p}_i \bar{x}_i \left(\hat{p}_i + \hat{x}_i + \frac{1}{2}(\hat{p}_i + \hat{x}_i)^2 \right) \right] + \frac{1}{2}h''(\bar{y}) \left[\sum_{i=1}^{2} \bar{p}_i \bar{x}_i (\hat{p}_i + \hat{x}_i) \right]^2. \tag{68}$$

Each self's optimality condition is given by[29]

[28] I drop constants which are irrelevant for utility maximization.

[29] I use $\bar{x}_i^{-\kappa_i} = \bar{p}_i h'(\bar{y})$.

$$-\kappa_i \hat{x}_i^* (s_i) = \hat{p}_i + \kappa_y E_i \left[\sum_{i=1}^{2} \frac{\mu_i}{\mu_y} \left(\hat{p}_i + \hat{x}_i^* (s_i) \right) \right], \tag{69}$$

which is the same as (19).

Substituting into the decision maker's utility in (68), taking an unconditional expectation of the objective, and using law of iterated expectation, the endogenous narrow thinking problem in (25) is equivalent to

$$\max_{\lambda_{-i,i}} -\frac{1}{2}\bar{\eta}w \left\{ \sum_{i=1}^{2} \mu_i \left(\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} + \kappa_y \left(\sum_{j=1}^{2} \frac{\mu_j}{\mu_y} \frac{\partial \hat{x}_j^{\text{Narrow}}}{\partial \hat{p}_i} \right) \right) \right\} \sigma_{p_i}^2 - \frac{1}{2} \sum_{i=1}^{2} \phi log_2 \left(\frac{1}{1-\lambda_{-i,i}} \right),$$

where $\bar{\eta} = h'(\bar{y}) = \frac{\partial v_i(\bar{x}_i)}{\partial x_i}/\bar{p}_i$ captures the marginal value of money at the point of approximation and I use the fact that all $\boldsymbol{\epsilon}$ and $\boldsymbol{p}$ are independent of each other and $I(\boldsymbol{s_{-i}}; \vec{\boldsymbol{p}}) = \frac{1}{2} log_2 \left(\frac{1}{1-\lambda_{-i,i}} \right)$. Now using the formulas of $\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i}$ and $\frac{\partial \hat{x}_{-i}^{\text{Narrow}}}{\partial \hat{p}_i}$ in (30) and (31), the endogenous narrow thinking problem is equivalent to

$$\min_{\lambda_{-i,i}} \frac{1}{2}\bar{\eta}w \left\{ \sum_{i=1}^{2} \frac{\left(\frac{\mu_i}{\kappa_i}\right)^2 (1-\kappa_i)^2 \sigma_{p_i}^2}{\frac{\mu_i}{\kappa_i} + \frac{\mu_{-i}}{\kappa_{-i}} \frac{\frac{\mu_y}{\kappa_y}\lambda_{-i,i}}{\frac{\mu_y}{\kappa_y} + (1-\lambda_{-i,i})\frac{\mu_{-i}}{\kappa_{-i}}} + \frac{\mu_y}{\kappa_y}} \right\} + \frac{1}{2}\phi \sum_{i=1}^{2} log_2 \left(\frac{1}{1-\lambda_{-i,i}} \right).$$

When ϕ is large enough, the above objective is convex. We can then use FOC to characterize the optimal $\lambda_{-i,i}^*$:[30]

$$\frac{\bar{\eta}w \left(\frac{\mu_i}{\kappa_i}\right)^2 (1-\kappa_i)^2 \sigma_{p_i}^2 \frac{\mu_{-i}}{\kappa_{-i}} \frac{\mu_y}{\kappa_y} \left(\frac{\mu_y}{\kappa_y} + \frac{\mu_{-i}}{\kappa_{-i}} \right)}{\left(\frac{\mu_i}{\kappa_i} \left(\frac{\mu_y}{\kappa_y} + \left(1-\lambda_{-i,i}^*\right) \frac{\mu_{-i}}{\kappa_{-i}} \right) + \left(\frac{\mu_y}{\kappa_y}\right)^2 + \frac{\mu_y}{\kappa_y}\frac{\mu_{-i}}{\kappa_{-i}} \right)^2} = \frac{\phi}{\ln 2 \left(1-\lambda_{-i,i}^*\right)}. \tag{70}$$

Let us define $f\left(\lambda_{-,i}^*\right) = \frac{\left(a\left(1-\lambda_{-i,i}^*\right)+b\right)^2}{\left(1-\lambda_{-i,i}^*\right)}$, with

$$a = \frac{\mu_1}{\kappa_1}\frac{\mu_2}{\kappa_2}$$

and

$$b = \left(\frac{\mu_y}{\kappa_y}\right) \left(\frac{\mu_1}{\kappa_1} + \frac{\mu_2}{\kappa_2} + \frac{\mu_y}{\kappa_y} \right).$$

We know that $f\left(\lambda_{j,i}^*\right)$ increases in $\lambda_{j,i}^* \in [0,1)$ if $a\left(1-\lambda_{-i,i}^*\right) < b$, which will be true if $\left(\frac{\mu_y}{\kappa_y}\right)\left(\frac{\mu_1}{\kappa_1} + \frac{\mu_2}{\kappa_2} + \frac{\mu_y}{\kappa_y}\right) > \frac{\mu_1}{\kappa_1}\frac{\mu_2}{\kappa_2}$. Together with the FOC in (70), the comparative statics in Proposition 16 follow directly.

The General Case: Flexible Information Acquisition

Environment. In the above example, I restrict each potential $\boldsymbol{s_i} \in \boldsymbol{\Omega_i}$ to have a particular form: each $\boldsymbol{s_i}$ consists of N noisy signals, one for each price. This is consistent with the information structure studied

[30] It is possible that, when ϕ is very large, an interior solution does not exist and $\lambda_{-i,i}^* = 0$. Even considering this case, Proposition 16 still holds with "weakly" increasing.

in the rest of the paper. An alternative is to let the potential signals depend on the fundamental flexibly. The only restriction on potential signals is that $\boldsymbol{s_1}, \boldsymbol{s_2}, \cdots, \boldsymbol{s_N}$ are always conditionally independent given $\vec{\boldsymbol{\theta}}$. That is, the noise in each decision i's signal about $\vec{\boldsymbol{\theta}}$ is idiosyncratic.[31] With such a flexible form of information acquisition, I can achieve a sharp characterization about the optimum of (25) 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 $\vec{\boldsymbol{x}}$ (and independent from $\vec{\boldsymbol{\theta}}$). Such terms will only add a constant to each optimal decision rule and will not influence the optimal information choice for each self.

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

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

Lemma 6 *With unrestricted $\boldsymbol{\Omega_i}$, in the optimum of (25), each decision i is based on an one-dimensional signal $\boldsymbol{s_i^*}$:*[32]

$$\boldsymbol{s_i^*} = \boldsymbol{\vartheta_i} + \mathbb{E}\left[\sum_{j\neq i} \gamma_{i,j} x_j^*\left(\boldsymbol{s_j^*}\right) | \vec{\boldsymbol{\theta}}\right] + \boldsymbol{\epsilon_i} \equiv \boldsymbol{t_i} + \boldsymbol{\epsilon_i}, \tag{71}$$

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

In (71), $\boldsymbol{\epsilon_i} \sim N\left(0, \sigma_i^2\right)$ 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 (71). Without the cognitive cost, the optimal decision i should be given by $\boldsymbol{\vartheta_i} + \sum_{j\neq i}\gamma_{i,j}x_j^*\left(\boldsymbol{s_j^*}\right)$. Now, with the cognitive cost, Lemma 6 shows that the optimal information for decision i will be given by a signal about the fundamental $\vec{\boldsymbol{\theta}}$ that is closest to $\boldsymbol{\vartheta_i} + \sum_{j\neq i}\gamma_{i,j}x_j^*\left(\boldsymbol{s_j^*}\right)$. The variance of the noise in this signal is pinned down by decision i's allocated cognitive cost 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.

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

[32] 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.

Given the optimal signal in (71), one can then solve optimal decision rules $\{x_i^*(\cdot)\}_{i=1}^N$. From (71), we know each self's optimal signal in turn depends on other selves' optimal decision rules. Solving this fixed-point problem, one can then characterize how each optimal signal $\boldsymbol{s_i^*}$ depends on the fundamental $\vec{\boldsymbol{\theta}}$.

Proposition 17 *The optimal signals depend on the fundamental* $\vec{\boldsymbol{\theta}}$ *as follows:*

$$\begin{pmatrix} E\left[\boldsymbol{s_1^*}|\vec{\boldsymbol{\theta}}\right] \\ \\ \cdots \\ \\ E\left[\boldsymbol{s_N^*}|\vec{\boldsymbol{\theta}}\right] \end{pmatrix} = \begin{pmatrix} \boldsymbol{t_1} \\ \\ \cdots \\ \\ \boldsymbol{t_N} \end{pmatrix} = \left(\mathbb{I}_N - \begin{pmatrix} 1 & & \cdots & \lambda_{N-1}^* & \lambda_N^* \\ \lambda_1^* & 1 & \cdots & \lambda_{N-1}^* & \lambda_N^* \\ & & \cdots & & \\ \lambda_1^* & \lambda_2^* & \cdots & 1 & \lambda_N^* \\ \lambda_1^* & \lambda_2^* & \cdots & \lambda_{N-1}^* & 1 \end{pmatrix} \circ \Gamma \right)^{-1} \begin{pmatrix} \boldsymbol{\vartheta_1} \\ \cdots \\ \boldsymbol{\vartheta_k} \\ \cdots \\ \boldsymbol{\vartheta_N} \end{pmatrix}, \tag{72}$$

where $\Gamma = \{\gamma_{i,j}\}_{1\leq i,j\leq N}$ *and* $\gamma_{i,j} = -\frac{\partial^2 v}{\partial x_i^2} / \frac{\partial^2 v}{\partial x_i \partial x_j}$. $\lambda_i^* = \frac{\sigma_{t_i}^2}{\sigma_i^2+\sigma_{t_i}^2} = 1 - 2^{-2\tau_i^*} \in (0,1)$ *is pinned down by decision* i'*s allocated cognitive cost* τ_i^*.

Similar to Proposition 10, 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 problem in (25) with two decisions. From (72), 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 cost is severe (ϕ is large 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.

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

Proposition 18 *In the optimum of (25),*

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

Proposition 18 shows that more volatile decisions (with high $Var(\boldsymbol{t_i})$) 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 in bitcoins (volatile $\boldsymbol{t_i}$) than to the self who invests in ETFs (stable $\boldsymbol{t_i}$).

Proof of Lemma 6. A necessary condition for $\{\boldsymbol{s_i^*}, x_i^*(\cdot)\}_{i=1}^N$ to be an optimum of (25) is that, for each i, $(\boldsymbol{s_i^*}, x_i^*(\cdot))$ is optimally chosen, taking the allocated cognitive cost τ_i^* and other $\left\{\boldsymbol{s_j^*}, x_j^*(\cdot)\right\}_{j\neq i}$ as given. That is, $(\boldsymbol{s_i^*}, x_i^*(\cdot))$ solves

$$\max_{\boldsymbol{s_i}\in\boldsymbol{\Omega_i}, x_i(\cdot)} E\left[u\left(x_1^*(s_1^*), \cdots, x_i(s_i)\cdots, x_N^*(s_N^*), \vec{\theta}\right)\right] \tag{73}$$

$$s.t. \quad I\left(\boldsymbol{s_i}; \vec{\boldsymbol{\theta}}\right) \leq \tau_i^* \tag{74}$$

As u is quadratic, maximizes the objective in (73) is equivalent to maximizing

$$E\left[\frac{u_{i,i}}{2}\left(x_i(s_i) - \vartheta_i - \sum_{j\neq i}\gamma_{i,j}E\left[x_j^*\left(s_j^*\right)|\vec{\theta}\right]\right)^2 + f\left(\left\{x_j^*\left(s_j^*\right)\right\}_{j\neq i}, \vec{\theta}\right)\right], \tag{75}$$

where $u_{i,i} = \frac{\partial^2 u}{\partial x_i^2}$ and I use the fact that $\boldsymbol{s_1}, \boldsymbol{s_2}, \cdots, \boldsymbol{s_N}$ is conditionally independent given $\vec{\boldsymbol{\theta}}$. The problem based on the objective in (75) and the constraint in (74) is then the standard tracking problem with quadratic loss function and Normally distributed target. From Sims (2003), we know the optimal signal $\boldsymbol{s_i}$ takes the form in Lemma 6.

Proof of Proposition 17. Given the chosen information $\{\boldsymbol{s_i^*}\}_{i=1}^N$, the optimal decision rule $\{x_i^*(\cdot)\}_{i=1}^N$ can be characterized by

$$x_i^*(\boldsymbol{s_i^*}) = E\left[\boldsymbol{\vartheta_i} + \sum_{j\neq i}\gamma_{i,j}x_j^*\left(\boldsymbol{s_j^*}\right)|\boldsymbol{s_i^*}\right] = \lambda_i^*\boldsymbol{s_i^*},$$

where $\lambda_i^* = \frac{\sigma_{t_i}^2}{\sigma_i^2+\sigma_{t_i}^2}$. Together with (71), we have

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

This leads to (72).

Proof of Proposition 18. 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

$$\begin{pmatrix} \boldsymbol{t_1} \\ \\ \cdots \\ \\ \boldsymbol{t_N} \end{pmatrix} = \left(\mathbb{I}_N - \begin{pmatrix} 1 & \Lambda_2 & \cdots & \Lambda_{N-1} & \Lambda_N \\ \Lambda_1 & 1 & \cdots & \Lambda_{N-1} & \Lambda_N \\ & & \cdots & & \\ \Lambda_1 & \Lambda_2 & \cdots & 1 & \Lambda_N \\ \Lambda_1 & \Lambda_2 & \cdots & \Lambda_{N-1} & 1 \end{pmatrix} \circ \Gamma\right)^{-1} \begin{pmatrix} \boldsymbol{\vartheta_1} \\ \cdots \\ \boldsymbol{\vartheta_k} \\ \cdots \\ \boldsymbol{\vartheta_N} \end{pmatrix}. \tag{76}$$

Based on Lemma 6 and Proposition 17, in the optimum of (25), $\left\{\lambda_i^*, \sigma_i^2\right\}_{i=1}^N$ defined in Lemma 6 and Proposition 17 must solve

$$\max_{\left\{\Lambda_i, \varsigma_i^2\right\}_{i=1}^N} g\left(\left\{\Lambda_i, \varsigma_i^2\right\}_{i=1}^N\right) - \phi h\left(\left\{\varsigma_i^2\right\}_{i=1}^N\right) \tag{77}$$

where $h\left(\left\{\varsigma_i^2\right\}_{i=1}^N\right) = \frac{1}{2}\sum_i \log_2 \frac{\varsigma_i^2 + \sigma_{t_i}^2}{\varsigma_i^2}$ and $\sigma_{t_i}^2$ is the variance of $\boldsymbol{t_i}$ defined based on (76). Rewrite g in a way similar to (75), 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 $\left\{\lambda_i^*, \sigma_i^2\right\}_{i=1}^N$ must solve (77), in the optimum of (25), we must have

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

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 18 follows.

References

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

Angrist, Joshua, Sydnee Caldwell, and Jonathan Hall. (2021) “Uber vs. Taxi: A Driver’s Eye View.” *American Economic Journal: Applied Economics.*

Barberis, Nicholas, Ming Huang, and Richard Thaler. (2006) “Individual Preferences, Monetary Gambles, and Stock Market Participation: A case for Narrow Framing.” *American Economic Review*, 96(4): 1069–1090.

Benartzi, Shlomo, and Richard Thaler. (1995) “Myopic Loss Aversion and the Equity Premium Puzzle.” *The Quarterly Journal of Economics*, 110(1): 73–92.

Bergemann, Dirk, Tibor Heumann, and Stephen Morris. (2017) “Information and Interaction.” *mimeo.*

Farber, Henry. (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.

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.

Golub, Benjamin, and Stephen Morris. (2017) “Expectations, Networks, and Conventions.” *mimeo.*

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

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

Morris, Stephen, and Ming Yang. (2019) “Coordination and Continuous Choice.” *Duke mimeo.*

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.

Redelmeier, Donald, and Amos Tversky. (1992) “On the Framing of Multiple Prospects.” *Psychological Science*, 3(3): 191–193.

Samuelson, Paul. (1963) “Risk and Uncertainty: A Fallacy of Large Numbers.” *Scientia*, 108–113.

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

Tversky, Amos, and Daniel Kahneman. (1992) "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, 5(4): 297–323.

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