

CHOICE SIMPLIFICATION: A THEORY OF MENTAL BUDGETING AND NAIVE DIVERSIFICATION*

BOTOND KŐSZEGI AND FILIP MATĚJKA

We develop a theory of how an agent makes basic multiproduct consumption decisions in the presence of taste, consumption opportunity, and price shocks that are costly to attend to. We establish that the agent often simplifies her choices by restricting attention to a few important considerations, which depend on the decision at hand and affect her consumption patterns in specific ways. If the agent's problem is to choose the consumption levels of many goods with different degrees of substitutability, then she may create mental budgets for more substitutable products (e.g., entertainment). In some situations, it is optimal to specify budgets in terms of consumption quantities, but when most products have an abundance of substitutes, specifying budgets in terms of nominal spending tends to be optimal. If the goods are complements, in contrast, then the agent may—consistent with naive diversification—choose a fixed, unconsidered mix of products. And if the agent's problem is to choose one of multiple products to fulfill a given consumption need (e.g., for gasoline or a bed), then it is often optimal for her to allocate a fixed sum for the need. *JEL* Codes: D01, D11, D14

I. INTRODUCTION

Individuals and households must make myriad decisions on how to allocate money in the face of many competing uses and a barrage of relevant information. A central part of [Thaler's \(1985, 1999\)](#) influential framework of mental accounting proposes that to help solve such allocation problems, individuals create different “mental budgets” for different purposes (entertainment, clothing, etc.), and treat these budgets as separate when responding to changes in circumstances. Yet despite the intuitive appeal of and empirical support for the concept, there is no theory that explains how a person creates separate mental budgets from fungible finances, and how this process interacts with her reactions to shocks.

*Formerly titled “An Attention-Based Theory of Mental Accounting.” We thank Yuriy Gorodnichenko, Paul Heidhues, Rupal Kamdar, Marc Kaufmann, David Laibson, Andrei Shleifer, Philipp Strack, four anonymous referees, and seminar audiences for excellent comments. This project has received funding from the European Research Council (grant agreement numbers 678081 and 788918) and the UNCE project (UNCE/HUM/035).

© The Author(s) 2020. Published by Oxford University Press on behalf of President and Fellows of Harvard College. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

The Quarterly Journal of Economics (2020), 1153–1207. doi:10.1093/qje/qjz043.
Advance Access publication on January 3, 2020.

In this article, we formulate a theory of expenditure allocation based on the premise that a person's attention is costly as well as flexible, so that she is motivated to both economize on attention and direct it toward important issues. We show that as a result, the person often engages in choice simplification: she restricts attention to a few considerations customized to be most useful for the decision at hand. The way she simplifies in turn affects her consumption patterns in specific ways, allowing us to explain mental budgeting, identify a connection between mental budgeting and naive diversification—a phenomenon that has hitherto been treated separately—and make other predictions.

After illustrating the logic of our results in a simple example in [Section II](#), in [Section III](#) we develop tools for analyzing the effects of costly attention on decision making when an agent's action and information are multidimensional. We extend the rational-inattention approach of [Sims \(2003\)](#) using the water-filling algorithm from information theory ([Telatar 1999](#); [Cover and Thomas 2006](#)) to show that the agent establishes a pecking order of information vectors and tilts her attention toward the more important vectors, potentially even ignoring the least important ones. Beyond consumption problems, our general methods are likely to apply to many economic situations, such as how individuals digest complex information about the economy or form political opinions.

In [Section IV.A](#), we turn to our main topic, consumption decisions with costly attention. We analyze how a person allocates expenditure when she faces independently distributed shocks to her preferences or consumption opportunities for (but not the prices of) different products, and she can reduce any aspect of that uncertainty through costly attention. We assume that the goods can be grouped into nested consumption categories and, first considering the case of substitutes, we posit that they are more substitutable within than between categories. For instance, a restaurant dinner and a play could both be in the "entertainment" category under the larger category of "discretionary spending," with the two being more substitutable with each other than either is with products outside the entertainment category.

Our main result says that the agent often behaves as if she had separate mental budgets for separate categories: (i) consumption in a category is independent of shocks to other categories, and (ii) total consumption is unresponsive, but individual consumption levels are smoothly responsive, to shocks within the category. In a classical consumption problem, (i) holds only if

utility is separable across categories—which our model does not assume—and (ii) does not hold for any utility function we could think of. Intuitively, the most relevant consideration for the agent to think about is which of multiple highly substitutable products are worth buying, so if she has sufficiently costly attention, she simplifies her decision making by thinking only about this consideration. As a result, she does not think about shocks to the optimal level of consumption, and hence her budget is fixed. Even if her attention cost is lower and therefore she does not have a hard budget, her spending in a category varies less than with full information, so she can be interpreted as having a soft budget.

Our budgeting result helps explain evidence that many individuals and households separate expenditures into budgetary categories (Rainwater et al. 1959; Kahneman and Tversky 1984; Lave 1995; Ameriks, Caplin, and Leahy 2003; Antonides, De Groot, and Van Raaij 2011), and makes the novel prediction that products are grouped into mental budgets according to their substitutability. Through a simple reinterpretation, our theory predicts that individuals may use budgeting strategies for other types of decisions, for instance, allocating separate time budgets for substitute tasks.

We illustrate that mental budgeting can interact in an economically interesting way with budget constraints. Even more than an unconstrained agent, a budget-constrained agent may prefer not to think about how much to consume in total, leading her to mentally budget. Furthermore, if her budget constraint is relatively tight, her mental budget exhausts all of her available funds—despite lower consumption being optimal with some probability. For budget-constrained individuals, therefore, costly attention increases consumption as well as the marginal propensity to consume out of increases in available funds.

An entirely different prediction emerges when we assume that the products are complements, and (paralleling the case of substitutes) they are more complementary within than between categories. Because the optimal consumption levels of complementary products tend to move together, the agent may now simplify her choice by not thinking about her relative values for products at all, only about how much she should consume in total. Hence, she may choose a fixed, unconsidered mix of products. Furthermore, such an unconsidered mix can also be optimal for substitute products if the agent's preferences for the products are sufficiently positively correlated. We argue that these predictions

are consistent with the phenomenon of naive diversification in financial (Benartzi and Thaler 2001, 2007) and consumption (Simonson 1990) decisions. Suppose, for instance, that the products are funds in an employer-based retirement program, and they look *ex ante* identical to the agent (e.g., because she knows nothing about her values for the individual funds to start with). If the funds invest in different assets, then they are complements because buying them jointly serves diversification purposes; if the funds invest in similar assets, then the agent's values for them are highly correlated. In either case, the agent may follow the $\frac{1}{N}$ rule, investing equal amounts in the available funds. Mental budgeting and naive diversification can therefore be viewed as solutions to the same type of decision-making problem that apply in different circumstances.

In Section V, we ask whether the agent still wants to set budgets for substitute products when there are price shocks, and whether she prefers budgets expressed in quantities of consumption or amounts of spending. Accordingly, we allow the agent to make and execute plans in two different ways: she can choose the quantity of consumption for each product, or she can choose the amount of spending on each product. Although these two ways of thinking are equivalent in a classical consumer problem in which prices are known, in our framework and with price uncertainty—in which the agent may not fully learn prices before making decisions—they are not equivalent. We establish that thinking in terms of spending is optimal whenever optimal total consumption is sufficiently price sensitive, or there are sufficiently many substitute products in a category. Intuitively, fixing the amount to be spent on a product means that consumption responds to unforeseen changes in the product's price, and this is optimal if an average of the relevant optimal price elasticities (of substitution and total consumption) is sufficiently high. Furthermore, we show that under reasonable conditions, a consumer who thinks in terms of spending sets spending budgets. These results explain the prevalence of spending budgets as well as the greater prevalence of spending budgets among (generally more price-sensitive) lower-income households, but they also predict consumption budgets in some plausible circumstances. For instance, a high-income consumer may set an entertainment budget in terms of nights out per month.

In Section VI, we consider a variant of our model in which the agent has unit demand for each product—for example, she

needs a single mattress or computer to replace her old one or a given amount of gasoline to drive that month—but has multiple versions of each product to choose from. Similar to the above, we ask whether deciding the version of the product (e.g., the grade of gasoline) to buy or the amount to spend is optimal for a consumer who does not process all price information before making decisions. We show that thinking in terms of spending is optimal if and only if product prices are on average sufficiently positively correlated with premiums for better products. If prices and premiums are uncorrelated, then it is optimal for the agent to absorb a price increase fully by increasing spending. If prices and premiums are highly correlated, however, an increase in prices greatly increases the average marginal price of increasing quality, so it is better to lower spending back to the original level. Thinking in terms of spending implies, in line with evidence by [Hastings and Shapiro \(2013\)](#) on gasoline purchases, that when prices for all varieties of a product rise, the agent switches to a lower-priced variety. Although in Hastings and Shapiro's specific setting the price and price premium are not positively correlated, our explanation applies if such situations are sufficiently uncommon in consumers' lives and consumers do not think about the specific setting separately.

In [Section VII](#), we discuss how our model relates to existing theories. Whereas previous work explores another central aspect of mental budgeting, self-control problems ([Shefrin and Thaler 1988](#); [Galperti 2019](#)), we are the first to explain how a person creates mental budgets from fungible finances and the first to formally connect mental budgeting and naive diversification. In [Section VIII](#), we note that our model does not distinguish between, and therefore cannot explain differential consumption responses to, different but fully fungible sources of income. We argue, however, that closely related attention-based models can potentially explain some of these phenomena. In addition, we emphasize that it would be fruitful to study the interaction between our framework and related phenomena, especially self-control problems and loss aversion.

II. EXAMPLE

In this section, we illustrate the logic of our main insights using a simple example.

II.A. Setup

The agent chooses the consumption levels of two goods, y_1 and y_2 , to maximize the expectation of

$$(1) \quad (\bar{x} + x_1)y_1 + (\bar{x} + x_2)y_2 - \frac{y_1^2}{2} - \frac{y_2^2}{2} - \theta y_1 y_2 - (y_1 + y_2),$$

where $\bar{x} > 1$ is her average taste for the goods, x_1 and x_2 are independent taste shocks drawn from $N(0, 1)$, and $\theta \in (-1, 1)$ is a substitutability parameter, with the goods being substitutes for $\theta > 0$ and complements for $\theta < 0$. Both goods have a price equal to 1, and the disutility of spending \$1 is also 1, so $y_1 + y_2$ is the total disutility of spending.

Before choosing y_1 and y_2 , the agent can observe exactly one of x_1 , x_2 , $x_1 - x_2$, and $x_1 + x_2$: she can think about her taste for one of the goods or her relative or total taste for the two goods. We ask: what does she optimally choose to think about, and how does this affect her consumption?

II.B. Solution: Mental Budgeting versus Naive Diversification

To facilitate an answer, we put the problem in a different form. Instead of working with the tastes x_1 and x_2 , we work with the relative and total tastes, $x_- = x_1 - x_2$ and $x_+ = x_1 + x_2$; and instead of solving for the consumption levels y_1 and y_2 , we solve for the relative and total consumption levels, $y_- = y_1 - y_2$ and $y_+ = y_1 + y_2$. Up to additive terms that are functions of x_1 , x_2 , and \bar{x} only—which the agent cannot influence—the objective expression (1) can then be written as

$$(2) \quad -\frac{(x_- - (1 - \theta)y_-)^2}{4(1 - \theta)} - \frac{(x_+ + 2(\bar{x} - 1) - (1 + \theta)y_+)^2}{4(1 + \theta)}.$$

To maximize her expected utility conditional on her information, the agent therefore chooses

$$(3) \quad y_- = \frac{E[x_- | \text{info}]}{1 - \theta} \quad \text{and} \quad y_+ = \frac{2(\bar{x} - 1) + E[x_+ | \text{info}]}{1 + \theta}.$$

Plugging the optimal y_- and y_+ into equation (2) yields that in choosing what to observe, the agent aims to maximize the

expectation of

$$-\frac{(x_- - E[x_- | \text{info}])^2}{4(1 - \theta)} - \frac{(x_+ - E[x_+ | \text{info}])^2}{4(1 + \theta)},$$

conditional on her information, which is $\frac{1}{4}$ times

$$(4) \quad -\frac{1}{1 - \theta} \cdot \text{var}[x_- | \text{info}] - \frac{1}{1 + \theta} \cdot \text{var}[x_+ | \text{info}].$$

Optimal information acquisition is now obvious from how information affects the conditional variances of x_- and x_+ . If the products are substitutes (i.e., $\theta > 0$, and therefore $\frac{1}{1-\theta} > \frac{1}{1+\theta}$), the agent chooses to observe x_- because she wants to know her relative taste for the two products. Since x_- and x_+ are independent, observing x_- provides no information about x_+ , so $y_+ = \frac{2(\bar{x}-1)+E[x_+]}{1+\theta} = \frac{2(\bar{x}-1)}{1+\theta}$. This means that $y_1 + y_2$ is constant: the agent has a fixed budget determined by her average taste \bar{x} for the products. Because $y_- = \frac{x_-}{1-\theta}$, however, the consumption levels y_1 and y_2 are not fixed—the agent does respond to changes in circumstances, but not by changing her total budget.

If the products are complements (i.e., $\theta < 0$, and therefore $\frac{1}{1-\theta} < \frac{1}{1+\theta}$), then the agent chooses to observe x_+ —she wants to know her total taste for the products. As a result, she learns nothing about x_- , so $y_- = \frac{E[x_-]}{1-\theta} = 0$. This means that $y_1 = y_2$: the agent naively diversifies, always choosing the goods in equal proportion. Since $y_+ = \frac{2(\bar{x}-1)+x_+}{1+\theta}$, however, the consumption levels y_1 and y_2 are not fixed—the agent does think about the problem, but not by changing the ratio in which she buys the products.

II.C. Other Implications

We substantially generalize the insights above and derive other predictions in Section IV.A, and study the implications of price uncertainty in Section V. We use variants of our simple model to make a few further points, which we do not reconsider in a more general setting. First, our model assumes that x_1 and x_2 are independent. If x_1 and x_2 are positively correlated, then $\text{var}[x_+] > \text{var}[x_-]$, which by equation (4) increases the value of observing x_+ . Hence, in this case naive diversification is more likely to occur. Intuitively, if the tastes for two products are highly positively correlated, the consumer is unlikely to learn much from thinking about which one she likes, so she chooses to think about

her total taste. Conversely, a negative correlation between x_1 and x_2 increases the value of observing x_- , increasing the tendency toward budgeting.

Second, by treating the disutility of spending money (or, equivalently, the value of saving) as a constant, we have implicitly assumed that the agent knows it or does not want to think about it. Uncertainty in the value of saving affects the disutility of spending on both products equally, so—if the agent can lower the uncertainty through thinking—it is equivalent to a positive correlation between x_1 and x_2 .¹ If the value of saving is highly uncertain, therefore, the previous point implies that our budgeting result fails. In this sense, figuring out one's value of saving to a point where one no longer wants to think about it much is a precursor to budgeting.

Third, consider also what happens when the goods are substitutes, and the agent has a relatively tight budget constraint $y_+ \leq y_+^{\max} \leq \frac{2(\bar{x}-1)}{1+\theta}$. Without information, the constraint would be binding, with the agent choosing $y_+ = y_+^{\max}$ and $y_- = 0$. Because observing x_- is only useful for choosing y_- , the constraint, which does not restrict y_- , leaves the value of observing x_- unchanged. In contrast, since observing x_+ , x_1 , or x_2 is useful for choosing y_+ , the constraint—which prevents increases in y_+ in response to news—decreases the value of observing any of these variables. Hence, the agent still prefers to observe x_- , and her total consumption is $y_+ = y_+^{\max}$, that is, she always exhausts her spendable funds. Intuitively, while consuming less might be optimal, thinking about this is less valuable than thinking about how to split her spendable funds between the goods. When the budget constraint is relatively tight, therefore, limited attention increases consumption.

Furthermore, notice that the agent spends a marginal increase in available funds if and only if her optimal unconstrained consumption exceeds available funds. Hence, her average marginal propensity to consume out of increases in available funds equals the probability with which her budget constraint

1. Formally, let the disutility of spending be $1 + \mu$, with μ capturing the uncertainty in the value of money, and let x'_1 and x'_2 be the independent taste shocks. The agent's utility is then $(\bar{x} + x'_1)y_1 + (\bar{x} + x'_2)y_2 - \frac{y_1^2}{2} - \frac{y_2^2}{2} - \theta y_1 y_2 - (1 + \mu)y_1 - (1 + \mu)y_2$. Setting $x_1 = x'_1 - \mu$ and $x_2 = x'_2 - \mu$ gives expression (1), where x_1 and x_2 are now positively correlated.

binds. As a result, limited attention also increases the average marginal propensity to consume from as low as $\frac{1}{2}$ to 1.²

Fourth, budget constraints can interact with attention costs. To illustrate, we suppose that acquiring a second signal from the set $\{x_1, x_2, x_-, x_+\}$ has a positive rather than infinite marginal cost. Once the agent knows x_- , any other signal perfectly reveals x_1 and x_2 , so the other signals are equally valuable. Because the budget constraint does not affect the value of learning only x_- but lowers the value of learning both x_1 and x_2 , it can lead the agent not to acquire a second signal. In this case, the budget constraint induces budgeting, as well as the associated focus on relative tastes, due to costly attention.

III. THEORETICAL TOOLS

In this section, we develop a methodology for analyzing rational-inattention models in which—as with mental budgeting and naive diversification—the agent’s information and action are multidimensional.³ Because these tools are potentially applicable to many economic settings, we present them in a general form. We lay out our results on consumption decisions in a self-contained way, so readers not interested in the general tools can skip to [Section IV.A](#).

III.A. Multidimensional Rational Inattention

The agent maximizes the expectation of the utility function $U(\mathbf{y}, \mathbf{x})$, which depends on an exogenous random vector of states $\mathbf{x} \in \mathbb{R}^J$ and her chosen vector of actions $\mathbf{y} \in \mathbb{R}^N$, less the cost of

2. To see that with full attention the marginal propensity to consume can be as low as $\frac{1}{2}$, suppose that $y_+^{\max} = \frac{2(\bar{x}-1)}{1+\theta}$, and note that optimal unconstrained total consumption is $y_+ = \frac{2(\bar{x}-1)+x_+}{1+\theta}$. Since x_+ has mean zero, the constraint binds with probability one-half.

3. For some previous applications of rational inattention, see [Veldkamp \(2006\)](#); [Mackowiak and Wiederholt \(2009\)](#); [Woodford \(2009\)](#); [Luo and Young \(2014\)](#); [Caplin and Dean \(2015\)](#); [Matějka and McKay \(2015\)](#); and [Matějka \(2016\)](#). See [Mackowiak, Matějka, and Wiederholt \(2018\)](#) for a review. Recent papers by [Afrouzi and Yang \(2019\)](#), [Fulton \(2017\)](#), [Miao, Wu, and Young \(2019\)](#), and [Verstyuk \(2019\)](#) also solve models of multidimensional rational inattention. These papers demonstrate that the agent may prefer lower-dimensional signals but do not identify implications for consumption patterns.

information processing. U takes the form

$$U(\mathbf{y}, \mathbf{x}) = -\mathbf{y}'C\mathbf{y} + \mathbf{x}'B\mathbf{y},$$

where $B \in \mathbb{R}^{J \times N}$, $C \in \mathbb{R}^{N \times N}$, and C is symmetric and positive definite. The matrix C summarizes interactions between actions, whereas B summarizes interactions between states and actions. We assume that the prior uncertainty about \mathbf{x} is multivariate Gaussian with the variance-covariance matrix ψ . To focus on the allocation of attention driven by preferences only, we let $\psi = \sigma_0^2 I$.

Before choosing \mathbf{y} , the agent can obtain any Gaussian signal about \mathbf{x} . The resulting posterior beliefs are also Gaussian, with the agent being able to choose the posterior variance-covariance matrix Σ subject to the constraint that $\psi - \Sigma$ is positive definite—that is, that the posterior is more precise than the prior. Denoting by $|\cdot|$ the determinant of a matrix, we posit that the cost of information is $\frac{\lambda}{2} \cdot (\log |\psi| - \log |\Sigma|)$, where $\lambda \geq 0$ is the agent's attention cost. This specification of decision making with costly attention is the reduced form of a general rational-inattention model in which the agent (instead of being restricted to a Gaussian signal) can choose any signal at a cost equal to the reduction in the entropy of her beliefs.⁴

As in the previous literature, there are three main reasons for using the entropy-based functional form for attention costs. First, it is highly tractable. Second, it has the basic property that information is costly (if the agent learns \mathbf{x} more precisely, then $|\Sigma|$ is lower, and therefore $\frac{\lambda}{2} \cdot (\log |\psi| - \log |\Sigma|)$ is higher). Third, it implies that all information has the same cost—what matters is the amount of uncertainty reduction, not what the uncertainty is about—so it can be viewed as ideal for studying information acquisition based on endogenous considerations about the benefits of information, not based on exogenous assumptions about the costs of information.

At the same time, researchers have raised various concerns about specifying attention costs to be linear in entropy reduction. [Woodford \(2012\)](#) points out that the entropy-based cost function fails to predict the finding from perceptual experiments that

4. [Sims \(2003\)](#) shows that in a multidimensional rational-inattention model with entropy costs, it is optimal for an agent with our linear-quadratic consumption utility to collect Gaussian signals; hence, we simply assume that the agent does so. In addition, the entropy of a Gaussian distribution with variance-covariance matrix Σ is a constant plus $\frac{\log |\Sigma|}{2}$.

subjects make smaller errors in more likely states. [Dean and Neligh \(2017\)](#) find that experimental subjects' behavior is consistent with a cost function that is convex in entropy reduction. Similarly, [Morris and Strack \(2017\)](#) establish that a constant marginal cost of signals in sequential information-acquisition problems corresponds to a convex entropy-based cost function. Accordingly, theoretical work generalizes the entropy-based cost function to allow for differences in comparison costs across versus within nests of products ([Fosgerau, Melo, and Shum 2017](#)), on different dimensions of the state space ([Pomatto, Strack, and Tamuz 2019](#)), and for nearby versus distant states ([Morris and Yang 2016](#)), and [Caplin and Dean \(2015\)](#) study a broader class of cost functions called posterior separable. With alternatives going beyond entropy-based costs, our decision problem would be difficult or impossible to analyze. Such extensions would add the consideration that the agent is more likely to obtain less costly information, but there is no obvious sense in which they might undermine the logic of our results.

III.B. Optimal Information Acquisition and Actions

Our method for solving the model, which we detail in the proof of Proposition 1 in [Appendix A](#), is analogous to the water-filling algorithm in the engineering literature (see, e.g., [Telatar 1999](#)). We first show that the agent's objective, expected utility less the cost of information, can be written as

$$(5) \quad -E[(\tilde{\mathbf{x}} - \mathbf{x})' \Omega (\tilde{\mathbf{x}} - \mathbf{x})] + \frac{\lambda}{2} \log |\Sigma|,$$

where $\Omega = \frac{BC^{-1}B'}{4}$ and $\tilde{\mathbf{x}}$ is the random mean of the posterior beliefs about \mathbf{x} , which depends on the realization of noise in signals. The first term in [expression \(5\)](#) is the expected loss from misperceptions ($\tilde{\mathbf{x}} - \mathbf{x}$), which are distributed according to $N(0, \Sigma)$ and translated into losses by Ω . The second term is the cost of information, with the constant $\frac{\lambda}{2} \cdot \log |\psi|$ dropped.

1. Decomposition into One-Dimensional Problems. Let v^1, \dots, v^J be an orthonormal basis of eigenvectors of the loss matrix Ω (which is symmetric), with the eigenvalue corresponding to v^i denoted by Λ_i . The utility term in [expression \(5\)](#) can be conveniently expressed using the transformation of coordinates to this

basis. Letting $(\tilde{\mathbf{x}} - \mathbf{x}) = \sum_i \tilde{\eta}_i v^i$, we have

$$(\tilde{\mathbf{x}} - \mathbf{x})' \Omega (\tilde{\mathbf{x}} - \mathbf{x}) = \left(\sum_i \tilde{\eta}_i v^i \right)' \Omega \left(\sum_i \tilde{\eta}_i v^i \right) = \sum_i \Lambda_i \tilde{\eta}_i^2.$$

The eigenvalue Λ_i is thus a scaling parameter for how uncertainty about the linear combination $(v^i \cdot \mathbf{x})$ translates into losses. Now the expectation of $\tilde{\eta}_i^2$ is by definition the posterior variance of $v^i \cdot (\tilde{\mathbf{x}} - \mathbf{x})$. Because the x_i are i.i.d. with prior variance σ_0^2 , the random variables $(v^i \cdot \mathbf{x})$ are also i.i.d. with prior variance σ_0^2 . Let us denote the posterior variance of $(v^i \cdot \mathbf{x})$ by $\sigma_i^2 \leq \sigma_0^2$. In the proof we show that Σ must be diagonal in the basis of the eigenvectors, and thus $\log |\Sigma| = \sum_i \log \sigma_i^2$. The agent's problem therefore reduces to

$$(6) \quad \max_{\sigma_i^2 \leq \sigma_0^2} \left(\sum_i -\Lambda_i \sigma_i^2 + \frac{\lambda \log \sigma_i^2}{2} \right).$$

This can now be solved separately for each i , yielding a simple information-acquisition strategy.

PROPOSITION 1 (INFORMATION ACQUISITION). The optimal information-acquisition strategy is to acquire independent signals of $v^i \cdot \mathbf{x}$ such that the posterior variance of $v^i \cdot \mathbf{x}$ is $\min\{\sigma_0^2, \frac{\lambda}{2\Lambda_i}\}$.

Intuitively, the agent processes more information about vectors in the space of \mathbf{x} that are more costly to misestimate. Specifically, if $\sigma_0^2 \leq \frac{\lambda}{2\Lambda_i}$, then the agent acquires no information about $v^i \cdot \mathbf{x}$; and if $\sigma_0^2 > \frac{\lambda}{2\Lambda_i}$, then she observes a signal about $v^i \cdot \mathbf{x}$ with precision chosen to bring the posterior variance of $v^i \cdot \mathbf{x}$ down to $\frac{\lambda}{2\Lambda_i}$. Hence, when the cost of information λ is high ($\frac{\lambda}{2\Lambda_i} > \sigma_0^2$ for all i), then the agent does not process any information. If the cost is somewhat lower, then the agent processes information about the $v^i \cdot \mathbf{x}$ with the highest Λ_i , but she processes no other information. At even lower costs, the agent processes information about $v^i \cdot \mathbf{x}$ for more i 's, and so on.

2. Responsiveness of Actions. Next we discuss implications for actions. We show in the Appendix that $\mathbf{y} = H\tilde{\mathbf{x}}$, where $H = \frac{C^{-1}B'}{2}$. We define ϕ_i^λ as the average change in the action \mathbf{y} when

\mathbf{x} changes in direction v^i by 1. We can think of φ_i^λ as the average responsiveness of the agent's behavior to shocks along v^i . Using this notation, the responsiveness under perfect information—when the agent has no attention costs—is φ_i^0 .

PROPOSITION 2 (OPTIMAL ACTIONS).

- (i) The space of actions is spanned by $\{Hv^i \mid \frac{\lambda}{2\Lambda_i} < \sigma_0^2\}$.
- (ii) The agent underresponds to shocks relative to the perfect-information case ($\varphi_i^\lambda < \varphi_i^0$), with

$$(7) \quad \frac{\varphi_i^0 - \varphi_i^\lambda}{\varphi_i^0} = \min\left(1, \frac{\lambda}{2\sigma_0^2 \Lambda_i}\right).$$

- (iii) In the range $\Lambda_i > \Lambda_j > \frac{\lambda}{2\sigma_0^2}$, the relative responsiveness $\frac{\varphi_i^\lambda}{\varphi_j^\lambda}$ is strictly increasing in λ .

Part (i), an immediate implication of Proposition 1, says that the agent's action moves only in response to information that is sufficiently important to pay attention to. Part (ii) says that the agent underresponds to shocks. Because the agent pays only partial attention to information, on average she does not notice the extent of shocks, so she does not respond as much as an agent with zero attention costs. More interestingly, part (iii) says that with costly attention, optimal behavior calls for concentrating reactions to shocks in directions that are the most important. As a result, the responsiveness to shocks along v^i relative to v^j is higher than under perfect information if and only if $\Lambda_i > \Lambda_j$.

IV. CONSUMPTION PATTERNS

We now apply the tools from Section III to analyze how a person attends and responds to taste or consumption opportunity shocks when choosing a consumption basket from many products with different degrees of substitutability or complementarity. This generalizes our example in Section II by allowing for more products and by assuming that the consumer can choose any type and any amount of information (rather than one of four signals). We analyze price shocks in the next section.

There are N goods, each of which has a price equal to 1. The agent's utility from consumption or, equivalently, spending levels

$y_1, \dots, y_N \in \mathbb{R}$ takes the quadratic form

$$(8) \quad - \sum_m y_m^2 - \sum_{m \neq n} \Theta_{mn} y_m y_n + \sum_m (\bar{x}_m + x_m) y_m - \sum_m y_m,$$

where $\Theta \in \mathbb{R}^N \times \mathbb{R}^N$ with $\Theta_{mm} = 1$ is a symmetric positive definite matrix that captures the substitutability patterns between the goods, \bar{x}_m is the baseline marginal utility of consuming good m , and x_m is a shock to this marginal utility. Randomness in one's marginal utilities could arise from uncertainty about taste—the agent does not know what combination of restaurant dinners, shirts, housing amenities, and so on maximizes her well-being—or from shocks to consumption opportunities—for example, if better bands happen to be in town, then the marginal utility of going to concerts is higher. Finally, $\sum_m y_m$ is the disutility of spending money.

To be able to analytically solve and economically interpret our model, we posit the following specific structure for Θ . The goods can be grouped into $L \geq 1$ level of nested categories. The level $l = L$ is the largest category (e.g., discretionary spending), which includes all N goods; the level $l = L - 1$ is the set of second-largest categories (e.g., entertainment), and so on, with the smallest ($l = 1$) categories being individual consumption goods (e.g., a dinner out). We denote by $R^{k,l} \subset \{1, \dots, N\}$ the consumption category k at level l . We assume that all categories at level l are of the same size ($|R^{k,l}| = |R^{k',l}|$ for all k, k', l), and that each category at level $l < L$ is a subset of a higher category (for each $l < L, k$, there is a unique k' such that $R^{k,l} \subset R^{k',l+1}$). The substitutability of two goods is determined by the smallest category to which they both belong. For two goods m and n , let l be the smallest l' such that there is a k with $m, n \in R^{k,l'}$. Then, $\Theta_{mn} = \gamma^l$, where γ^2 through γ^L are constants.⁵

We assume that the x_m are i.i.d. normal random variables with mean 0 and variance σ_0^2 , and the agent can obtain any multivariate normal signal about (x_1, \dots, x_N) . Part of this thinking could, for instance, involve mentally simulating future consumption (as in [Gabaix and Laibson 2017](#)) or searching for information about consumption opportunities. The agent's cost of attention

5. Note that we have introduced the notion of categories merely to facilitate the definition of the substitutability matrix Θ and the statement of our results; we do not presume that the agent thinks of goods in the same category separately from other goods.

is the same as in Section III, so that she maximizes the sum of her expected utility given her posterior beliefs plus $\frac{\lambda \log |\Sigma|}{2}$, where $\lambda \geq 0$ is her cost of attention, Σ is the variance-covariance matrix of her posterior, and $|\Sigma|$ is the determinant of Σ .

The attention cost in our model is most straightforwardly interpreted as an information-acquisition or information-processing cost. But it can also be interpreted as a calculation cost when the agent knows her tastes and consumption opportunities (or, in Section V, prices), but without thinking does not know what they imply for optimal consumption. Under either interpretation, the assumption that the agent can think about the vector (x_1, \dots, x_N) in a fully flexible way is unrealistic. For instance, it is unlikely that one can obtain a noisy signal about an arbitrary linear combination of the marginal utilities of this month's entertainment programs. At the same time, there is clearly flexibility in what a person thinks about or focuses on, and our framework captures such flexibility without making potentially ad hoc assumptions on its limits. Fortunately, the optimal ways of thinking we identify below are highly plausible and intuitive, so if we allowed only plausible ways of thinking, the same solutions would obtain.⁶

As a benchmark, we identify how the agent behaves if she has costless attention and how she responds to ex ante known changes (i.e., changes in the \bar{x}_m). For instance, the agent's average taste may evolve over time. To state the result, let $\mathbf{y} = (y_1, \dots, y_N)'$, $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_N)'$, $\mathbf{x} = (x_1, \dots, x_N)'$.

FACT 1. If $\lambda = 0$, then $\mathbf{y} = \frac{\Theta^{-1}(\bar{\mathbf{x}} + \mathbf{x})}{2}$. For any $\lambda \geq 0$, $E[\mathbf{y}] = \frac{\Theta^{-1}\bar{\mathbf{x}}}{2}$.

The agent's average behavior responds to ex ante known changes in exactly the same way as with perfect information. This also means that her utility function (i.e., the matrix Θ) can be extracted from her responses to ex ante known changes. As we show later, her responses to ex post shocks she needs to think about are often markedly different and by implication do not accurately reflect her true preferences over consumption. Nevertheless, these responses

6. Relatedly, our formal framework, in which the agent solves a complex optimization problem, may suggest a view of consumers going through a conscious, elaborate, and precise thought process to arrive at their budgets and strategies to spend. As with many other models in microeconomic theory, a more realistic interpretation is that consumers approximate the solution through trial and error or other means and make a habit of strategies that have worked in the past.

can be predicted from her (from ex ante known shocks measurable) true preferences.

IV.A. *Substitutes: Mental Budgeting*

First, we consider substitute goods, assuming that $0 < \gamma^L < \dots < \gamma^2 < 1$. This captures the idea that a good is a better substitute for other goods in its category than for goods in a different category. For instance, a French dinner is a better substitute for a Chinese dinner than for a movie. Then:

PROPOSITION 3 (HARD BUDGETING OF SUBSTITUTE PRODUCTS). If

$0 < \gamma^L < \dots < \gamma^2 < 1$, then there are $\lambda_1, \dots, \lambda_L$ satisfying $\lambda_L < \dots < \lambda_1$ such that

$$(9) \quad \lambda \geq \lambda_l \Leftrightarrow \sum_{m \in R^{k,l}} y_m = \text{constant for all } k.$$

Proposition 3 says that if (and only if) her attention cost is sufficiently high, the agent has a fixed mental budget—a constant total expenditure—for each l -category of products. Accordingly, the higher is her cost of attention—for example, because she has lower cognitive ability or is busy with other things—the more likely she is to budget, and the narrower are her budgets. To appreciate ways in which such behavior differs from that of a classical decision maker, suppose that $\lambda_2 < \lambda < \lambda_1$, and one category at level 2 is entertainment. Denoting the entertainment category by R :

COROLLARY 1. (i) For any $m \in R$ and $n \notin R$, y_m does not depend on x_n ; (ii) $\sum_{m \in R} y_m$ is constant; and (iii) for any $m \in R$, $E[y_m | \mathbf{x}]$ is a smooth function of the vector $(x_m - x_n)_{n \in R \setminus \{m\}}$ that is strictly increasing in each component.

Corollary 1 implies two related phenomena. First, part (i) says that the agent's consumption decisions regarding entertainment are independent of other shocks. In a classical consumption problem, this occurs only if the utility from entertainment is separable from the rest of the utility function. We do not impose such separability; in fact, with full information $\frac{\partial y_m}{\partial x_n} < 0$ for all $n \neq m$.⁷ Second, parts (ii) and (iii) imply that the agent's total consumption of entertainment is independent of shocks, but her consumption within

7. See Fact 2 in Appendix B for a proof.

the category responds smoothly to within-category shocks. This is not the case in any classical model we could think of.

Intuitively, knowing about a shock to the relative marginal utility of movies and theater is very valuable, as it allows for substantial readjustment of both consumption levels through substitution. Knowing about a shock to the relative marginal utility of movies and clothing is less valuable, because the scope for substitution between these goods is lower. And knowing about a shock to the marginal utility of movies is also less valuable, as it leads mainly to the adjustment of movies consumption. With the agent's attention being costly, she thinks only about the most important consideration, the relative marginal utility of movies and theater. As a result, she fixes total entertainment consumption.

Proposition 3 explains evidence that a significant number of consumers have category-specific budgets. As a stark manifestation of this phenomenon, many households used to place budgets allocated for different purposes into different envelopes or tin cans (Rainwater et al. 1959; Lave 1995). Furthermore, Ameriks, Caplin, and Leahy (2003) and Antonides, De Groot, and Van Raaij (2011) document that the mental budgeting (if not physical separation) of expenses is still common. More generally, the majority of U.S. households report having a spending plan or budget (e.g., Lin et al. 2016), which may also be broken down into smaller budgets. Indeed, most of the many online financial management tools seem to presume that users want to set separate budgets for separate categories. Proposition 3 not only accounts for this evidence but makes the novel prediction that the most substitutable goods go into the same budget.

Of course, testing this prediction requires identifying which products are most substitutable. One way to do so is to use the agent's responses to ex ante known information to measure Θ , as we mentioned after Fact 1. In addition, the proof of Proposition 3 implies that ex post information is useful as well: although the agent does not maximize her consumption utility, she does substitute more between product pairs m, n for which $\theta_{m,n}$ is greater. Formally, for two products m and n in a budget, the smaller is the smallest common category to which both belong, the greater is the agent's average tendency to substitute between the products in response to a relative taste shock, $\frac{\partial E[y_m - y_n | x_m - x_n]}{\partial (x_m - x_n)}$.⁸

8. See Corollary 3 in Appendix B for a proof.

This allows an observer to establish a ranking of product pairs according to substitutability. After observing a person's or population's substitutability ranking, it is possible to predict, for instance, how a comparable person or population with greater attention cost might form budgets.

The logic of Proposition 3 applies in other domains as well. For instance, there is evidence that some individuals have mental budgets for time allocation, such as hours per day devoted to studying (Rajagopal and Rha 2009). This follows from our model by reinterpreting y_m as the time allocated to task m , and x_m as a shock to the return of working on task m . Furthermore, our theory predicts that a person creates budgets for substitute tasks, for instance, different ways of studying for an exam.

Our model is static in the sense that the agent solves a single optimization problem over what information to obtain and what to consume. But she does not have to make all choices at the same time. When choosing budgets, she can leave her plans incomplete and obtain information about shocks only when relevant consumption opportunities start arising, even making decisions separately for separate categories of products. This piecemeal execution is facilitated by the separable nature of the optimal plan, and is in fact optimal if obtaining the same information or mentally simulating consumption at the earlier budgeting stage is costlier.⁹

Having budgets leads to specific patterns in how a person reacts to shocks. Suppose, for instance, that the x_m in the entertainment category all increase by the same amount—that is, unusually fun entertainment opportunities present themselves across the board. Then the agent's average consumption of entertainment as well as other goods remains unchanged. Because she evaluates entertainment goods only relative to each other, on average she does not see a reason to change her behavior. If she had unlimited attention, in contrast, she would respond to such positive shocks by increasing entertainment consumption

9. Technically speaking, at the budgeting stage it is necessary for the agent to understand exactly what she will do at the execution stage. Interpreted more broadly, it is sufficient for her to have (perhaps based on experience) a reasonable understanding of the average value of increasing her budget.

Relatedly, when the agent acquires information piecemeal, the question arises how costly each piece of information is. A simple assumption consistent with our formulation is that at each stage, the cost of information equals $\frac{\lambda}{2}(\log|\Sigma_0| - \log|\Sigma_1|)$, where Σ_0 and Σ_1 are the variance-covariance matrices of her previous and new beliefs, respectively.

and decreasing other consumption. Similarly, if a single x_m increases, that leads the agent to increase y_m . If she had full information, she would also decrease the consumption of all other goods. Because she has a budget, however, she concentrates the substitution to within the category.

For simplicity our setup imposes a strong form of symmetry on products, but a simple extension of our proof makes it clear that weaker assumptions can also lead to budgeting. For instance, suppose that $L = 3$, there are two categories of unequal size at level 2, and products are symmetric within each category. Then, for sufficiently high attention costs the agent has a budget for each category. More generally, for an agent with high attention costs to budget within a category, it is sufficient for utility within the category to have the symmetric nested structure we have imposed; other categories could have different structures.

Proposition 3 identifies a particularly strong form of budgeting, in which the budget is completely fixed: if $\lambda \geq \lambda_l$, then the correlation between the consumption of a good and the total consumption of other goods in its l -category is -1 . Although this stark hard-budgeting result does not hold in many other cases, it identifies a force toward budgeting that holds more generally. As a case in point, we identify a version for lower attention costs.

PROPOSITION 4 (SOFT BUDGETING OF SUBSTITUTE PRODUCTS). Suppose that $0 < \gamma^L < \dots < \gamma^2 < 1$ and $\lambda < \lambda_l$. For any k and $m \in R^{k,l}$, the correlation between y_m and $\sum_{n \in R^{k,l} \setminus \{m\}} y_n$ is strictly decreasing in λ .

Proposition 4 says that the higher the agent's attention cost, the more she restricts consumption adjustments to substitutions within a category. In this sense, she can be viewed as having a soft budget for l -categories.

Asymmetries in the prior variances of x_m or prices also lead to a kind of soft budgeting. To illustrate, suppose that $L = 2$ and $N = 4$ —there is a single category of four products—and the prior variances $\sigma_{0,m}^2$ satisfy $\sigma_{0,1}^2 \neq \sigma_{0,2}^2 = \sigma_{0,3}^2 = \sigma_{0,4}^2$. We show in Appendix B (Proposition 9) that there are λ_1 and α such that if $\lambda \geq \lambda_1$, then $\alpha y_1 + y_2 + y_3 + y_4$ is constant, with $\alpha < 1$ if and only if $\sigma_{0,1}^2$ is greater than the other $\sigma_{0,m}^2$. Hence, total spending equals $y_1 + y_2 + y_3 + y_4 = \text{constant} + (1 - \alpha)y_1$. Furthermore, simulations show that unless the asymmetry is very large, an increase in y_1 is associated with a decrease in $y_2 + y_3 + y_4$ much more than with

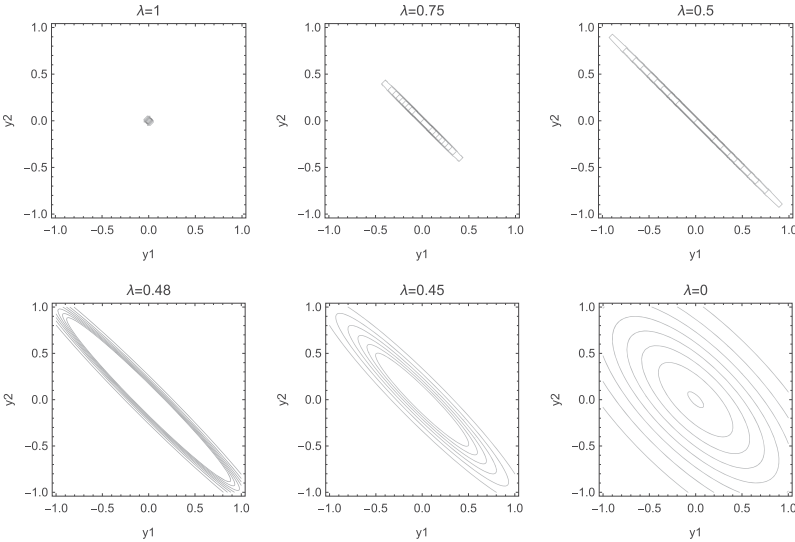


FIGURE I

Example

Joint distributions of y_1 and y_2 for different costs of attention when $N = 4$, there are two categories, $\{1, 2\}$ and $\{3, 4\}$, and $\sigma_0^2 = 1, \gamma^2 = \frac{1}{2}, \gamma^3 = \frac{1}{4}$. Iso-density curves are shown.

full information. This can be interpreted as saying that the agent has a soft target budget, allowing herself to go over the target if she happens to have a high value for a good with more volatile value. Relatedly, if good 1 has price $p_1 \neq 1$, then total spending is $p_1 y_1 + y_2 + y_3 + y_4 = \text{constant} + (p_1 - \alpha)y_1$: now the agent allows herself to go over the target if she has a high value for a more expensive product. When choosing between cheaper chicken and more expensive beef, for instance, she allows herself to splurge when especially nice beef is available.

Figure I illustrates Propositions 3 and 4 in an example. We consider four goods grouped into categories $\{1, 2\}$ and $\{3, 4\}$, and draw the joint distribution of y_1 and y_2 for different levels of λ . For costless attention ($\lambda = 0$), the distribution of possible consumption pairs is quite dispersed. At the other extreme, for a very high attention cost ($\lambda = 1$), the consumption amounts are fixed. For lower, but relatively high attention costs ($\lambda = 0.75, 0.5$), the agent sets a budget for the two products, so her consumption is always on the same budget line. These situations correspond to

Proposition 3. For even lower positive attention costs ($\lambda = 0.48, 0.45$), the agent starts substituting goods 1 and 2 with goods 3 and 4, but not as much as with costless attention, so the distribution of y_1 and y_2 is closer to a budget line than for $\lambda = 0$. These situations correspond to Proposition 4.

IV.B. Complements: Naive Diversification

We turn to complementary products, assuming that $\gamma^2 < \dots < \gamma^L < 0$. This means that products are arranged in a nested fashion into categories and are stronger complements within than across categories. For instance, different features of a car (e.g., driving experience, seats, sound system) might be highly complementary to each other but not to one's furniture. To simplify our statement and capture situations in which the products are ex ante equally desirable, we also assume that the \bar{x}_m are equal. Then:

PROPOSITION 5 (NAIVE DIVERSIFICATION). If $\gamma^2 < \dots < \gamma^L < 0$ and the \bar{x}_m are equal, then there are $\lambda_2, \dots, \lambda_L$ satisfying $\lambda_2 < \dots < \lambda_L$ such that

$$(10) \quad \lambda \geq \lambda_l \Leftrightarrow \text{for any } k \text{ and any } m, n \in R^{k,l}, y_m = y_n.$$

Proposition 5 says that if the agent's attention cost is sufficiently high, then she engages in naive diversification, that is, chooses a fixed mix of products, in category l . Intuitively, because the optimal consumption levels of complementary products tend to move together, the agent does not think about their optimal relative consumption at all, only about how much she should consume in total. Continuing with the example of cars, the agent does not think separately about the quality of the engine, seats, sound system, and so on, she wants—she only thinks about whether she wants an economy or luxury car.

Although it is more typical for complementary products, under some circumstances naive diversification also occurs for substitutes. First, as we have illustrated in Section II, it emerges if the x_m are sufficiently positively correlated. Second, it emerges—in a trivial form—if the agent's attention cost is so high that she does not obtain any information. In this case, her consumption of all products is fixed at the ex ante optimal level, and therefore the mix of products is fixed as well.

An important application of the above results is naive diversification in financial decisions, whereby a person chooses a simple mix of investments that is unlikely to be fully optimal. For instance, [Benartzi and Thaler \(2001\)](#) document that many employees in employer-based retirement savings plans divide their investments equally across available funds, and relatedly, employees invest more in stocks if there are more stock funds available. [Huberman and Jiang \(2006\)](#) find a similar pattern for plans offering 10 or fewer funds, although not for plans offering more funds. To see how our model can account for this phenomenon in an example, suppose that an investor with mean-variance preferences decides the amounts y_1 and y_2 to invest into two assets. There are two equally likely states, with asset 1's net return being $x_1 + 1$ in state 1 and $x_1 - 1$ in state 2, and asset 2's net return being $x_2 - 1$ in state 1 and $x_2 + 1$ in state 2. It is easy to check that the mean of the investor's wealth is $x_1 y_1 + x_2 y_2$ and the variance is $(y_1 - y_2)^2$, so the utility function can be written in the form of [expression \(8\)](#) with $\Theta_{12} = \gamma_2 = -1$. Hence, [Proposition 5](#) predicts that an investor with sufficiently costly attention splits her investment equally between the two assets. More generally, because diversification is desirable, different investments are often complements, so [Proposition 5](#) predicts that investors may diversify naively.¹⁰

Of course, it may be the case that different funds invest in similar assets, so combining them does not serve a diversification purpose, and therefore the funds are more appropriately viewed as substitutes. In this case, however, the investor's values for the funds are prone to be highly positively correlated, making naive diversification optimal again. Reinforcing this tendency is that preferences are prone to be positively correlated to start with: that one fund is a good investment reflects in part that employer-sponsored retirement savings is a good investment in general, and therefore other funds in the program are good investments as well.

Investigating a completely different domain, [Simonson \(1990\)](#) finds that individuals naively diversify when choosing items to

10. In the illustrative example above, the complementarity of the two investments relies on the asset returns being negatively correlated. Even for uncorrelated or somewhat positively correlated asset returns, investments are complements if the investor's disutility from variance is strictly concave. Furthermore, with a precautionary savings motive, risky and safe investments are often complements.

consume at different future dates.¹¹ Our model explains this finding if individuals have a taste for variety—which is equivalent to complementarity—and are subject to taste shocks. Consistent with our perspective, Simonson argues that naive diversification is attributable to the combination of taste uncertainty and the desire to simplify the decision.

The observation that the agent reacts to ex ante known changes exactly as in the full-information case (Fact 1) qualifies Proposition 5 in an interesting way. For instance, suppose that an investor distinguishes between stock funds and bond funds and knows that stocks are more valuable investments for her. Then she chooses more stock funds than bond funds or might choose only stock funds. But if she considers stock funds as ex ante identical, then she still naively diversifies within the class of stock funds. More generally, if the agent sees a reason to invest in only a handful of funds, but treats them as ex ante equally good investments, then she may naively diversify between these funds. Huberman and Jiang (2006) find some evidence of such a conditional $\frac{1}{N}$ rule.

For simplicity of presentation, we treat the cases of substitute products and complementary products separately. But it is easy to combine the two problems into one grand decision problem. In particular, suppose that a subset of the products are substitutes as above, while the rest are complements as above, with preferences over the two subsets being separable. Because the two problems are then separable, our results apply unchanged to each subset. Furthermore, although we have not investigated such cases in detail, more complicated patterns involving both budgeting and naive diversification of the same products can emerge if some categories are complements, and some are substitutes. Suppose, for instance, that dinners and movies, as well as jazz and cocktails, are complements, but they are substitutes across the two pairs. Then, an agent with high attention costs always pairs dinners and movies as well as jazz and cocktails—in this sense naively

11. In one study, for instance, students chose snacks to be received at the end of three different classes. When choosing the snacks one at a time at the beginning of these classes, 9% of students chose three different snacks. But when simultaneously choosing three snacks ahead of time, 64% of students chose three different snacks. To the extent that in the sequential-choice condition students know more about their momentary tastes, the former choices better reflect their true preferences.

diversifying—but has a budget for the total number of entertainment evenings.¹²

V. PRICE UNCERTAINTY AND THE NATURE OF BUDGETS

In this section, we establish a version of our budgeting result for situations characterized by price uncertainty and identify plausible conditions under which a mental budget is optimally set in terms of monetary spending rather than consumption.¹³ The agent has the same utility function as in Section IV.A, and for tractability we assume that $L = 2$ —there is a single category of substitute products—and the goods are symmetric. Furthermore, while the agent's tastes and consumption opportunities are deterministic, the prices of the goods, p_1 through p_N , are i.i.d. normally distributed random variables. Hence, the agent's consumption utility is

$$(11) \quad - \sum_m y_m^2 - \sum_{m \neq n} \theta y_m y_n + \sum_m \bar{x} y_m - \sum_m p_m y_m.$$

We assume that the agent can obtain information about the p_m in the same costly way as about the x_m in Section IV.A. As we have noted, an alternative interpretation of attention costs is re-optimization costs when the agent observes the price shocks, but must exert costly cognitive effort to figure out what these imply for optimal consumption.

We conceptualize the question of whether the agent might want a budget for consumption or for spending by asking a more fundamental question: whether she wants to think—that is, make plans and execute decisions—in terms of the consumption levels of the goods or the amounts of spending on the goods.¹⁴ Formally, in the former case she chooses consumption y_m for each good, and in the latter case she chooses spending $Y_m = p_m y_m$ on each good.

12. For a formal statement and proof, see Corollary 4 in Appendix B.

13. In our main application for naive diversification, retirement investment, decisions are naturally denominated in dollar amounts invested into funds. This corresponds to prices that equal 1, so there is no price uncertainty.

14. This type of question is almost never considered in the literature on rational inattention, but one notable exception is Reis (2006). Analyzing a consumption-savings problem in which a consumer does not know her wealth perfectly, Reis asks whether the consumer prefers to make decisions in terms of consumption or savings.

Although these two ways of thinking are equivalent in a classical problem with known prices, in our model—in which the agent does not fully learn prices before making decisions—they are not equivalent. For instance, deciding to buy a front-row ticket to a concert no matter how much it costs will generally not result in the same purchase as deciding to spend \$100 on the concert no matter where one sits.

The case in which the agent thinks in terms of consumption levels reduces to our previous analysis by setting $x_m = -p_m$, so our budgeting results from Section IV.A apply and mean that the agent sets consumption budgets. We now compare this to thinking in terms of spending. Denoting the means of p_m and y_m by \bar{p} and \bar{y} , respectively, and linearly approximating spending Y_m as $Y_m = (\bar{p} + (p_m - \bar{p}))(\bar{y} + (y_m - \bar{y})) \approx \bar{p}y_m + (p_m - \bar{p})\bar{y}$, we get

$$(12) \quad y_m \approx \frac{Y_m}{\bar{p}} - \frac{p_m - \bar{p}}{\bar{p}}\bar{y}.$$

To keep our model within the quadratic framework of Section III, we work with this approximation. The approximation retains a general property of thinking in terms of spending: that by fixing spending when she does not know the price, the agent makes the consumption level responsive to the unknown price. It is this general property, and not our use of an approximation, that drives the logic of Proposition 6.

To state our results, we define two measures of how the agent would respond to information if it was costless. Assuming for the definition that $\lambda = 0$, let

$$\epsilon^1 = \frac{\partial(y_m - y_n)}{\partial(p_m - p_n)} \frac{\bar{p}}{\bar{y}} \quad \text{and} \quad \epsilon^2 = \frac{\partial(\sum_m y_m)}{\partial(\sum_m p_m)} \frac{\bar{p}}{\bar{y}},$$

which are the optimal perfect-information elasticity of substitution between products and the optimal perfect-information elasticity of total consumption with respect to the total price, respectively.¹⁵ We find:

PROPOSITION 6. For any $\lambda, \sigma_0^2, \epsilon^1, \epsilon^2$, thinking in terms of spending yields strictly higher expected utility than thinking in

15. Because the utility function is quadratic, optimal perfect-information consumption is linear in prices. Hence, the above directional derivatives are constant. Furthermore, due to the symmetry of the problem, ϵ^1 does not depend on m and n .

terms of consumption if (i) $\epsilon^1, \epsilon^2 > \frac{1}{2}$ or (ii) $\epsilon^1 > \frac{1}{2}$ and N is sufficiently large, and the converse holds if (iii) $\epsilon^1, \epsilon^2 < \frac{1}{2}$.

Proposition 6 identifies two sufficient conditions for thinking in terms of nominal spending to be optimal. Both conditions require that the products are relatively good substitutes ($\epsilon^1 > \frac{1}{2}$). To understand the logic of condition (i), suppose first that $N = 1$, that is, there is a single product. Then, the condition says that the price elasticity of consumption of the single product must be greater than $\frac{1}{2}$. Intuitively, fixing nominal spending generates a price elasticity of consumption of 1 (from approximation (12), $\frac{(y_m - \bar{y})}{(p_m - \bar{p})} \frac{\bar{p}}{\bar{y}} = 1$) while fixing consumption generates a price elasticity of consumption of 0, so the former is optimal if and only if the optimal price elasticity is closer to 1 than to 0. Extending the logic to $N > 1$ gives condition (i): thinking in terms of spending is optimal if both relevant elasticities are greater than $\frac{1}{2}$. The converse gives condition (iii): thinking in terms of consumption is optimal if both relevant elasticities are less than $\frac{1}{2}$.

If $\epsilon^1 > \frac{1}{2}$ and $\epsilon^2 < \frac{1}{2}$, then the logic is not sufficient to determine whether thinking in terms of spending is optimal. Still, condition (ii) says that it is optimal if N is sufficiently large. Intuitively, this occurs because with many products, the predominant manner in which the agent wants to adjust consumption to shocks is by substituting between products—not by adjusting total consumption—so this substitution elasticity is more important in determining how she wants to think.

Our next proposition extends the budgeting result in Proposition 3 and Corollary 1 to spending.

PROPOSITION 7. Suppose that the agent thinks in terms of spending, and $\epsilon_1 \epsilon_2 > 1$. Then, there are λ_1, λ_2 satisfying $0 < \lambda_2 < \lambda_1$ such that if $\lambda_2 \leq \lambda < \lambda_1$, then total spending $\sum_m Y_m$ is constant, but the individual spending levels Y_m are not constant.

The logic also parallels that before: the most valuable pieces of information to know about are price differences, so often this is all the agent pays attention to. As a result, she restricts adjustments to substitutions between products, fixing total spending.¹⁶

16. The intuition for the qualifier $\epsilon_1 \epsilon_2 > 1$ derives from the central property of thinking in terms of spending, that it forces consumption to be sensitive to unanticipated price shocks. If the optimal (full-information) price elasticity of category

Thinking in terms of nominal spending, and having nominal budgets, is therefore optimal if the optimal price elasticity of total consumption is sufficiently high, or it is not too low and product categories feature many closely substitutable products. These results explain the general prevalence of spending budgets, and—since lower-income people have higher price elasticities of consumption—the greater prevalence of spending budgets among lower-income individuals.¹⁷ Nevertheless, our model predicts that individuals who do not care much about prices are more likely to have budgets expressed in terms of quantities. Consider, for example, a high-income person whose primary constraint in entertainment consumption is time, not money. Because she is therefore not price sensitive, she is more likely to choose a budget in entertainment quantity. Anecdotally, some people do seem to set consumption budgets, for instance, deciding to go out twice a month or to take two weeks of vacation a year. Relatedly, as [Krishnamurthy and Prokopec \(2010\)](#) note, in some self-control settings people tend to have quantity budgets, for instance in the number of weekly desserts or Weight Watchers points they allow themselves. Although our model does not formalize a self-control motive, this is another setting in which the primary cost of consumption is not the price, so that we predict quantity budgets rather than spending budgets.

Having a spending budget leads to an interesting pattern in how a person reacts to price shocks.

COROLLARY 2. Suppose that the agent thinks in terms of spending, and $\epsilon_1 \epsilon_2 > 1$, $\lambda_2 \leq \lambda < \lambda_1$. A decrease in the price of good m lowers spending on good m and increases spending on other goods.

consumption, ϵ_2 , is low, then it is important for the agent to pay attention to the price level to reduce unanticipated price shocks. Hence, in that case paying attention to the price level is more important than paying attention to price differences, so trading off only within the category is never optimal.

17. An additional potential reason for the last pattern is that lower-income individuals have higher costs of attention. Indeed, an experiment by [Mani et al. \(2013\)](#), and a variety of other evidence discussed in [Schilbach, Schofield, and Mullainathan \(2016\)](#), indicate that poverty impedes cognitive performance, which means that lower-income people have a higher λ . A classical account, however, would suggest that lower-income people have a lower opportunity cost of time due to lower wages, and therefore have a lower λ .

With full information, a decrease in the price of a good would lead to an increase in the consumption of that good and a decrease in the consumption of substitutes. In direct contrast, Corollary 2 says that the agent increases the consumption of substitutes as well. Although they are not precise confirmations, some experimental results are indicative of this prediction. In the experiment of Heilman, Nakamoto, and Rao (2002), shoppers who were given \$1 off an item increased their purchases of products related to the discounted item. But unlike in our model, the discount applied only to one item and hence was not a price decrease, and the discount also increased purchases of unrelated “treats.” Similarly, Heath and Soll (1996) find in hypothetical choices that MBA students reduce their entertainment consumption more if they had spent \$20 on a sports ticket than if they had received the same ticket as a gift. But again, a gift is not identical to a price shock.

Note that our model assumes linear disutility of money. Because thinking in terms of spending rather than quantities reduces risk in one’s total spending, a budget constraint over nominal spending, or more generally a concave utility function over nominal savings provides an additional reason to think in terms of spending and therefore to have spending budgets.¹⁸ Once again, this is especially likely to apply to low-income individuals, who typically face tighter constraints.

VI. UNIT DEMAND

In our main model, the agent chooses consumption levels from a continuum. In a number of prototypical consumer decisions, however, a person is better described as having unit demand, needing exactly one item and being able to choose it from a selection. For instance, in the medium run the car a person uses, and how much she uses her car, are fixed, so that she needs to buy a fixed amount of gasoline. When a consumer’s computer breaks, she needs to buy exactly one new computer to replace it.

18. The simplest formal way to make this point is to assume mean-variance preferences over spending. Start with our model above, in which the agent does not care about the variance of spending. Suppose that the agent wants to set budgets, and is indifferent between thinking in terms of spending and thinking in terms of consumption. Now suppose that she also derives disutility from the variance of her spending. Then, her achievable level of utility is strictly lower if she thinks in terms of quantities, but the same if she thinks in terms of spending. This means that she strictly prefers to think in terms of spending.

When shopping for a new bedroom, a homeowner may be looking for exactly one mattress and one comforter. We now analyze the implications of our framework for such purchases.

We assume that there are N categories of products. In each category, there is a continuum of products with different quality levels, and the agent is looking to buy exactly one of them (with her utility being $-\infty$ if there is a category in which she does not purchase). In category m , product $y_m \in \mathbb{R}$ has utility y_m , and a random price $p_m(y_m)$. For instance, one category m could be gasoline, with the choice of y_m representing the grade of gasoline that the agent chooses. The agent's total utility is $\sum_m (y_m - p_m(y_m))$.

The shape of the pricing functions $p_m(y_m)$ is determined by the differentiable, strictly increasing and strictly convex function $p(\cdot)$ that has full range and satisfies $\lim_{y \rightarrow -\infty} p'(y) < 1$ and $\lim_{y \rightarrow \infty} p'(y) > 1$. But the actual pricing functions are subject to shocks of the following form. For each category m , nature independently draws (i) the random variable x_m from a distribution with mean 0, and (ii) whether a vertical or a horizontal price shock occurs, which have probabilities s and $1 - s$, respectively. Then, the pricing function for category m becomes $p_m(y_m) = p(y_m) + x_m$ if the price shock is vertical, and $p_m(y_m) = p(y_m + x_m)$ if the price shock is horizontal. In combination with the assumption that $p(\cdot)$ is convex, this specification captures two canonical types of price shocks of interest for shopping behavior. A vertical price shock changes the price level while leaving the marginal price of increases in quality unchanged. With a horizontal price shock, however, a change in the price level changes the marginal price of increases in quality in the same direction; so (say) a price increase is associated with an increase in the marginal price as well.¹⁹

We consider an agent who has sufficiently costly attention (a sufficiently high λ) such that she does not want to think about price shocks, and therefore makes a plan that is independent of price realizations. An alternative interpretation is that the

19. In terms of consumption utility, our unit-demand model is an extremely simplified variant of our basic model: there is one product in each category with separable, linear utility. Unlike in our basic model, however, prices are now non-linear. An alternative approach would have been to assume that prices are linear and (to ensure interior solutions) that the utility function is concave. With this alternative specification, the condition for when fixing spending is optimal implicates both the pricing function and the utility function in a way that we do not find transparent. With our specification, the condition implicates only the pricing function and is transparent.

price uncertainty is the residual uncertainty after the agent has thought about the problem. Similarly to the previous section, we ask whether the agent wants to fix the level of quality or the amount of spending for each category. For computers, for instance, she could decide on a specific computer brand and configuration no matter how much it costs, or she could ask for the best \$2,000 computer no matter what specific machine that is. For gasoline, she could buy the same grade each time, or she could decide how much she is willing to spend on gas, and choose the grade whose price is closest to that amount. These choice variables seem equally easy to implement in practice: in the former case the agent needs to remember the version she wants to buy in each category, and in the latter case she needs to remember the price she is aiming for in each category.

PROPOSITION 8. Suppose that the agent does not acquire information about prices. For any $p(\cdot)$ and any mean-zero nondegenerate shock distribution, there is an $S \in (0, 1)$ such that fixing quality is optimal for $s > S$, and fixing spending is optimal for $s < S$.

If all price shocks are vertical ($s = 1$), then fixing quality is optimal. In this case, the marginal price of increasing y_m is constant, so choosing a fixed y_m is optimal. This means that if prices in category m increase, then the agent absorbs the shock and increases spending on category m . If some price shocks are horizontal, however, an increase in prices also increases the average marginal price of increasing y_m , so fully absorbing a price increase by increasing spending is not optimal. Instead, lowering spending back toward the original level increases utility. In the extreme case in which all price shocks are horizontal ($s = 0$), fixing spending is optimal. In this case, decreasing spending back to the original level perfectly aligns marginal value with marginal price. Extending this logic, fixing spending is superior to fixing quality if a sufficiently large share of price shocks is horizontal—or, equivalently, the price and marginal price of quality are sufficiently positively correlated.

Again, our model assumes linear disutility of money. If the agent has a budget constraint over nominal spending, or more generally her utility over nominal savings is concave, then she is more prone to think in terms of spending to reduce risk in her total spending.

The economically most important prediction of this section emerges when consumers choose to fix spending. Then, because $p_m(\cdot)$ is always strictly increasing, an increase in prices means that the agent must substitute to a lower-quality product. This prediction provides a potential explanation for the finding of [Hastings and Shapiro \(2013\)](#) that when gasoline prices rise, there is a shift in demand from premium to regular gasoline—that is, a cheaper product in the same category. Our explanation, however, requires that the circumstances that Hastings and Shapiro study are relatively rare in consumers' lives. In particular, Hastings and Shapiro find substitution toward lower-grade gasoline for price shifts for which the price and marginal price of quality are approximately uncorrelated, while our model predicts such behavior only if the price and marginal price of quality are sufficiently positively correlated. Nevertheless, the correlation between the price and marginal price of quality is plausibly positive for many product categories consumers have experience with. Even for gasoline, Hastings and Shapiro focus on the short run, and the correlation may be positive in the longer run. Hence, our model accounts naturally for the evidence under at least two circumstances, especially for budget-constrained consumers. First, if the relevant correlation is sufficiently positive for gasoline, then fixing spending is optimal even for gasoline purchases in isolation. Second, more plausibly, if the correlation is sufficiently positive for the average consumer product with unit demand, and the consumer does not want to think separately about the correlation or does not want a separate shopping strategy for gasoline, then again fixing spending on gasoline is optimal.

Note that for a consumer who prefers to think in terms of spending ($s < S$), the implications of our unit-demand model contrast in an interesting way with those of our continuous-demand model above. When the agent has a budget in the continuous model, an equal increase in prices—that is, a vertical price increase—for a category leaves her spending levels unchanged in expectation for all products. This occurs because she is paying attention only to the relative marginal utilities of products in the category, which a vertical price increase does not change. In the unit-demand model, in contrast, a vertical increase in prices leads the consumer to reallocate spending to a cheaper substitute.

VII. RELATED LITERATURE

Mental budgeting is a central component of, and is often referred to in the literature as, mental accounting, but the latter term is used for a broader set of issues. In applications of prospect theory, in particular, a mental account often refers to the set of monetary outcomes that are evaluated jointly in the context of a single decision (e.g., [Kahneman and Tversky 1984](#); [Thaler 1985](#); [Henderson and Peterson 1992](#)). For instance, a person is more willing to drive 20 minutes for a \$5 saving if it comes off of a \$15 purchase than if it comes off of a \$125 purchase ([Tversky and Kahneman 1981](#)), presumably because she evaluates the saving together with the purchase to which it is applied. Our article is instead about mental accounts/budgets that serve as a decision-making aid when there are multiple competing uses for money.

There are two main explanations for mental budgets that have long been noted in the literature. Our framework is close in spirit to one of them: the idea that mental budgets simplify a consumer's otherwise hopelessly complex problem by breaking it into manageable pieces ([Thaler 1999](#); [Zhang and Sussman 2018](#)). This idea, however, is very underdeveloped in the literature. Most notably, previous work does not formalize how mental budgets simplify decisions and does not provide precise predictions on how consumers group products into budgets. Indeed, the need for research on these questions seems widely recognized (e.g., [Hastings and Shapiro 2013](#); [Zhang and Sussman 2018](#)). We provide a theory of product categorization, and our formalization also generates other predictions, such as the connection we find between mental budgeting and naive diversification.

The other, more developed explanation for mental budgets is self-control problems—attempting to use budgets to mitigate overconsumption in the future. In a classic paper, [Shefrin and Thaler \(1988\)](#) develop a life cycle consumption-savings model in which the individual's "planner" self would like to control the "doer" self's tendency to consume too much. Shefrin and Thaler assume that the individual can separate money into different mental accounts, current spendable income, current assets, and future income, out of which her marginal propensities to consume are different. In the context of goal setting under self-control problems, [Koch and Nafziger \(2016\)](#) assume that a person can decide between broad and narrow goals, and that falling short of one's

chosen goal(s) leads to sensations of loss. The motive to avoid such losses creates an incentive that mitigates self-control problems.²⁰ Similarly, Galperti (2019) compares good-specific and total-expenditure budgets for a person who is subject to self-control problems and intratemporal and intertemporal taste shocks.

Our theory provides a complementary reason for mental budgets that has a different foundation and therefore different predictions and features. Conceptually, the most important difference is in the nature of mental budgets themselves: while self-control-based theories exogenously assume nonfungibility of money in the sense that spending from different accounts or budgets is subject to different constraints or preferences, in our model mental budgets emerge despite money being fully fungible. The different foundation also allows us to derive naive diversification from the same framework and make other predictions.

Gorman (1959) identifies circumstances under which it is optimal for a standard utility maximizer to make consumption decisions using a two-step procedure similar to that in Sections IV.A and V, whereby she first allocates fixed budgets to different consumption categories, and then optimizes within each category given the allocated budget. Unlike in our model, the budgeting in the first stage requires the agent to know with certainty all the relevant price indices for the categories, and there is no taste uncertainty.

In predicting that the agent may completely ignore some aspects of her decision environment, our model is similar to the sparsity-based model of bounded rationality by Gabaix (2014). In Gabaix's setting, the variables the agent may choose to look at are exogenously given, whereas in ours the agent can choose any combination of variables. We also apply the model to different questions than Gabaix.

Because our theory predicts unambiguous budgets based on economic preferences and fundamentals, it fails to capture some subtle context dependence in how individuals categorize outlays. For instance, Cheema and Soman (2006) find that individuals categorize a restaurant dinner flexibly as either food or entertainment depending on which budget has more money left over in it. The authors interpret such malleability in mental budgeting as an attempt to justify spending.

20. See also Hsiaw (2018) for related work and Pagel (2017) for other implications of loss aversion for consumption-savings behavior.

VIII. CONCLUSION

Although our models explain some important regularities in how individuals allocate money between multiple products, there are related phenomena that we have not covered. In particular, because our model does not distinguish multiple fully fungible sources of income, it cannot explain differences in whether and how consumers spend different types of income. The most important of these phenomena is the consumption effect of transfers that can only be used on a subset of products. The rational consumer model with full information implies that if such a transfer is *inframarginal*—that is, if the consumer would have spent more than the transfer on the products in question—then it is equivalent to cash. Yet experimental work by [Abeler and Marklein \(2016\)](#) and empirical work by [Hastings and Shapiro \(2018\)](#) indicate that *inframarginal* transfers have larger effects on the consumption of targeted products than do cash transfers. Even when a transfer is not *inframarginal*, it can have a surprisingly large effect: for instance, incentives for health-improving behaviors that are minute relative to the health benefits can significantly influence behavior ([Volpp et al. 2008](#); [Dupas 2014](#)).²¹ Although not predicted by our current framework, there is a plausible attention-based explanation for these findings. Namely, there are many things that a person could consider doing but that she deems not worthwhile to think about due to costly attention, and she therefore does not do them. Receiving a transfer or subsidy can induce the person to think about the potential benefits, increasing the effect of the transfer. In ongoing work, we formalize this mechanism and consider what it implies for the optimal design of transfers.

Of course, we do not believe that mental budgeting is solely about costly attention. As we have mentioned, a likely motive for creating mental budgets is self-control problems. It would be interesting to combine the attention-based and self-control-based explanations of mental budgeting to identify interactions. For example, a person may use the costly nature of her attention to improve self-control by creating plans that she is unwilling to reconsider later. And when it comes to implementing a mental-budgeting-based consumption plan, researchers understand

21. A related finding in political economy is the flypaper effect ([Hines and Thaler 1995](#)): when a local government receives a grant earmarked for a specific purpose, it tends to increase spending on that purpose by the amount of the grant.

that if the budget becomes a reference point, then loss aversion helps stick with the plan. Our theory provides one possible foundation for which outcomes are evaluated jointly in a reference-dependent model. Once again, it would seem fruitful to combine the attention-based view with loss aversion.

APPENDIX A: LQ MULTIVARIATE SETUP

Proof of Proposition 1. The quadratic utility function can be rewritten as

(13)

$$U(\mathbf{y}, \mathbf{x}) = -\left(\mathbf{y} - \frac{C^{-1}B'}{2}\mathbf{x}\right)' C \left(\mathbf{y} - \frac{C^{-1}B'}{2}\mathbf{x}\right) + \frac{\mathbf{x}'BC^{-1}B'\mathbf{x}}{4}.$$

If the posterior mean is $\tilde{\mathbf{x}}$, then the agent chooses an action (maximizing expected utility):

(14)
$$\mathbf{y} = \frac{C^{-1}B'}{2}\tilde{\mathbf{x}}.$$

This is because certainty equivalence applies in a quadratic setup. Plugging equation (14) into equation (13), the realized utility \tilde{U} for a state \mathbf{x} , but a posterior mean $\tilde{\mathbf{x}}$ is

(15)
$$\tilde{U}(\tilde{\mathbf{x}}, \mathbf{x}) = -(\tilde{\mathbf{x}} - \mathbf{x})'\Omega(\tilde{\mathbf{x}} - \mathbf{x}) + \mathbf{x}'\Omega\mathbf{x},$$

where $\Omega = \frac{BC^{-1}B'}{4}$. The first term is the loss from imperfect posterior beliefs, where $(\tilde{\mathbf{x}} - \mathbf{x})$ is the misperception. Given the variance-covariance matrix Σ for the distribution of $(\tilde{\mathbf{x}} - \mathbf{x})$, the expectation of the first term equals the trace of $\Omega\Sigma$. Because the second term in equation (15) depends on the realized state \mathbf{x} only, that is, it is independent of the agent's strategy, then the original problem takes the form:

(16)
$$\max_{\psi \geq \Sigma} -Tr(\Omega\Sigma) + \frac{\lambda}{2} \log |\Sigma|.$$

The second term in [expression \(16\)](#) is the cost of information, it is a log of the determinant of Σ .²² The larger the posterior uncertainty is, the lower the cost. The cost term here includes entropy of the posterior only, because entropy of a fixed prior amounts to an additive constant only. The condition $\psi \succeq \Sigma$ requires that $(\psi - \Sigma)$ is positive semidefinite, which means that acquisition of Gaussian signals cannot make beliefs less precise, that is, signals must have nonnegative precision.

To explore what signals the agent collects, let us decompose the loss matrix Ω , which is symmetric and thus has an orthonormal basis of eigenvectors. Let $\Omega = U\Lambda U'$, where U is a unitary matrix (the columns of which are eigenvectors of Ω), and Λ is a diagonal matrix with its elements Λ_{ii} equal to the eigenvalues Λ_i of Ω .

$$\begin{aligned}
 -Tr(\Omega\Sigma) + \frac{\lambda}{2} \log |\Sigma| &= -Tr(U\Lambda U'\Sigma) + \frac{\lambda}{2} \log |\Sigma| \\
 &= -Tr(\Lambda U'\Sigma U) + \frac{\lambda}{2} \log |U'\Sigma U| \\
 (17) \qquad \qquad \qquad &= -Tr(\Lambda S) + \frac{\lambda}{2} \log |S|,
 \end{aligned}$$

where $S = U'\Sigma U$ is the posterior variance-covariance matrix in the basis of eigenvectors of Ω . The condition $(\psi \succeq \Sigma)$ takes the form of $(U'\psi U \succeq S)$; note that $\psi = \sigma_0^2 I$.

Now we show by contradiction that S is diagonal. Suppose that the optimal S is not diagonal, and let S^D be the matrix constructed from its diagonal, that is, $S_{ii}^D = S_{ii}$ for all i and $S_{ij}^D = 0$ for all $i \neq j$.

First, since $\sigma_0^2 I - S$ is positive semidefinite, then $\sigma_0^2 I - S^D$ is also positive semidefinite. This is because for a diagonal S^D it suffices to check that $S_{ii}^D \leq \sigma_0^2$, which is implied by the fact that $\sigma_0^2 I - S$ is also positive semidefinite. Second, Hadamard's inequality implies:

$$(18) \qquad \qquad \qquad \frac{\lambda}{2} \log |S| \leq \frac{\lambda}{2} \sum_i \log S_{ii} = \frac{\lambda}{2} \log |S_D|,$$

22. Entropy of a multivariate $N(\mu, \Sigma)$ of dimension n is $\frac{n}{2}(\log(2\pi) + 1) + \frac{1}{2} \log |\Sigma|$.

where the equality holds if and only if S is diagonal. Third, $Tr(\Lambda S) = Tr(\Lambda S^D)$, because Λ is diagonal. Therefore, putting this together implies that S cannot be the optimum, since S^D delivers a higher objective due to the lower information cost, inequality (18), and is feasible.

Therefore, S is diagonal. Using equation (17), the original problem takes the form:

$$(19) \quad \max_{S_{ii} \leq \sigma_0^2} - \sum_{i=1}^N \left(S_{ii} \Lambda_i + \frac{\lambda \log(S_{ii})}{2} \right).$$

The first-order condition with respect to S_{ii} implies:

$$-\Lambda_i + \frac{\lambda}{2S_{ii}} = 0,$$

and the solution is

$$S_{ii} = \min \left(\sigma_0^2, \frac{\lambda}{2\Lambda_i} \right).$$

Proof of Proposition 2. Part (i): Proposition 1 implies that the space of posterior means $\tilde{\mathbf{x}}$ is spanned by all eigenvectors v^i for which $\frac{\lambda}{2\Lambda_i} < \sigma_0^2$, and the statement is then a trivial implication.

Part (ii): Let $\xi_i = 1 - \frac{S_{ii}}{\sigma_0^2}$ be the relative reduction of uncertainty about the component $v^i \cdot \mathbf{x}$. ξ is also the linear weight on a signal (as opposed to on the prior) in Bayesian updating with Gaussian signals. This means that in one-dimensional Bayesian updating, if the random variable $v^i \cdot \mathbf{x}$ moves by Δx , then the posterior mean about this variable moves in expectation by $\xi_i \Delta x$.

Because the agent chooses independent signals on $v^i \cdot \mathbf{x}$, Bayesian updating does in fact take the one-dimensional form. Responsiveness then is:

$$\varphi_i^\lambda = \frac{|H \xi_i v^i|}{|v_i|} = \frac{|\xi_i H v^i|}{|v_i|} = \xi_i \varphi_i^0.$$

This equation together with Proposition 1 implies the expression (7).

Part (iii): Differentiating $\frac{\varphi_i^\lambda}{\varphi_j^\lambda}$ with respect to λ then implies the statement. □

APPENDIX B: CONSUMPTION AND SPENDING BUDGETS

Proof of Fact 1. This is an immediate implication of equation (14) for $C = \Theta$ and $B = I$. □

Let r_l denote the size of the category $R^{k,l}$ on level l , that is, $r_l = |R^{k,l}|$ for all k .

LEMMA 1. Θ has a base of eigenvectors $\{v^{k,l,r'}\}_{l \in \{2..L\}, k \in \{1.. \frac{N}{l}\}, r' \in \{1..(r_l-1)\}}$, and $(1,..,1)$. $v^{k,l,r'}$ is associated with a category k of a level l , and it has the following properties

$$(20) \quad v_m^{k,l,r'} = 0 \quad \forall m \notin R^{k,l}$$

$$(21) \quad \sum_{m \in R^{k,l}} v_m^{k,l,r'} = 0$$

$$(22) \quad v_m^{k,l,r'} = v_n^{k',l,r'} \quad \forall m, n; \exists k' : m, n \in R^{k',l-1}.$$

Moreover, eigenvalues μ^l of $v^{k,l,r'}$ are given by:

$$\mu^l = \sum_{n \in R_m^{l-1}} (\Theta_{m,n} - \gamma^l) = \mu^{l-1} + (\gamma^{l-1} - \gamma^l)r_{l-1},$$

and $\mu^2 = \gamma^1 - \gamma^2$.

Proof of Lemma 1. Let us fix m , and apply Θ to an eigenvector associated with $R^{k,l}$; we drop the index m of the vector. Let R_m^{l-1} be the category on level $(l - 1)$ that the good m

belongs to.

$$\begin{aligned}
 \sum_n \Theta_{m,n} v_n^{k,l} &= \left(\sum_{n \notin R^{k,l}} \Theta_{m,n} v_n^{k,l} \right) + \left(\sum_{n \in R^{k,l} \setminus R_m^{l-1}} \Theta_{m,n} v_n^{k,l} \right) + \left(\sum_{n \in R_m^{l-1}} \Theta_{m,n} v_n^{k,l} \right) \\
 &= 0 + \gamma^l \left(\sum_{n \in R^{k,l} \setminus R_m^{l-1}} v_n^{k,l} \right) + \left(\sum_{n \in R_m^{l-1}} \Theta_{m,n} v_n^{k,l} \right) \\
 &= \gamma^l \left(\sum_{n \in R^{k,l}} v_n^{k,l} - \sum_{n \in R_m^{l-1}} v_n^{k,l} \right) + \left(\sum_{n \in R_m^{l-1}} \Theta_{m,n} v_n^{k,l} \right) \\
 &= \gamma^l \left(0 - \sum_{n \in R_m^{l-1}} v_m^{k,l} \right) + \left(\sum_{n \in R_m^{l-1}} \Theta_{m,n} v_m^{k,l} \right) \\
 (23) \quad &= v_m^{k,l} \sum_{n \in R_m^{l-1}} (\Theta_{m,n} - \gamma^l).
 \end{aligned}$$

The first equality is a simple decomposition into terms with elements within different categories. In the second, we used [equation \(20\)](#). The third is based on a decomposition of elements of $R^{k,l}$ into a subcategory with m and the other elements. The fourth equality uses [equation \(21\)](#) for the first term, and [equation \(22\)](#) is applied for the other two terms to substitute elements $v_j^{k,l}$ indexed by j by a constant $v_i^{k,l}$, since $v_j^{k,l}$ is constant in R_i^{l-1} .

Eigenvalue μ^l is therefore $\sum_{n \in R_m^{l-1}} (\Theta_{m,n} - \gamma^l)$. For $l = 2$ the only subcategory including m is m itself, $\mu^2 = \gamma^1 - \gamma^2$. And for $l > 2$:

$$\begin{aligned}
 \mu^l &= \mu^{l-1} + \gamma^{l-1} r_{l-2} - \gamma^l r_{l-1} - (r_{l-2} - r_{l-1}) \gamma^{l-1} \\
 (24) \quad &= \mu^{l-1} + (\gamma^{l-1} - \gamma^l) r_{l-1}.
 \end{aligned}$$

Therefore, each $v^{k,l,r'}$ is an eigenvector, and they form a basis. This is because they are all mutually orthogonal. Vectors associated with distinct categories are orthogonal due to [equation \(20\)](#), and vectors associated with mutually nested categories are orthogonal due to [equation \(22\)](#) and [equation \(21\)](#). For vectors of the same category, the dimensionality is due to [equation \(22\)](#) equal the number of subcategories minus one lower dimensionality due to [equation \(21\)](#), $\frac{r_l}{r_{l-1}} - 1$. The total number of vectors associated with level $l > 1$ is $\frac{N}{r_{l-1}} - \frac{N}{r_l}$, and the total number of these

orthogonal eigenvectors on all levels is $N - 1$, which together with the eigenvector $(1, \dots, 1)$ delivers N orthogonal eigenvectors, and thus a basis. \square

Proof of Proposition 3. We proceed in the following steps. First, we use Lemma 1 to infer the eigenvectors v^i and eigenvalues Λ^i of Ω . Second, we use Proposition 1 to find costs of information for which the agent acquires information about $v^i \cdot \mathbf{x}$. Third, we connect information acquisition to changes in $v^i \cdot \mathbf{y}$ by invoking Proposition 2. Finally, we show how this relates to fixed budgets.

First, because in this case $\Omega = \frac{\Theta^{-1}}{4}$, then Ω has the same eigenvectors v^i as Θ . For an eigenvector v^l associated with a level l , the eigenvalue is:

$$(25) \quad \Lambda_i = \frac{1}{4}(\mu^l)^{-1},$$

where μ^l is given recursively by equation (24). For nested substitutes, $\gamma^l < \gamma^{l-1}$, eigenvalues Λ_i are decreasing in the level l , since μ^l are increasing in l .

Second, let $\lambda_{l-1} = 2\sigma_0^2 \Lambda_i$. According to Proposition 1, the agent gets information about $v^l \cdot \mathbf{x}$ if and only if $\lambda < \lambda_{l-1}$. Note that λ_{l-1} are decreasing in l , because Λ^i are decreasing in l . Therefore, there are $\lambda_1, \dots, \lambda_L$ satisfying $\lambda_L < \dots < \lambda_1$ such that $v^i \cdot \tilde{\mathbf{x}}$ is constant for all v^i associated with a level l if and only if $\lambda \geq \lambda_{l-1}$.

Third, according to equation (14), $\mathbf{y} = \frac{\Theta^{-1}}{2} \tilde{\mathbf{x}}$. Because eigenvectors of $H = \frac{\Theta^{-1}}{2}$ are the same as those of Ω and Θ described above, then since $\{v^i\}$ form an orthogonal basis:

$$v^i \cdot \mathbf{y} = v^i \cdot H \tilde{\mathbf{x}} = v^i \cdot \sum_j (2\Lambda_j)(v^j \cdot \tilde{\mathbf{x}})v^j = 2\Lambda_i(v^i \cdot \tilde{\mathbf{x}}),$$

which is constant if and only if $v^i \cdot \tilde{\mathbf{x}}$ is constant, which holds according to step 2 above if and only if $\lambda \geq \lambda_{l-1}$.

Finally, we are interested in expressing $\sum_{m \in R^{k,l}} y_m$, which equals $(\mathbf{1}^{k,l} \cdot \mathbf{y})$, where $\mathbf{1}^{k,l}$ is an indicator vector of the category $R^{k,l}$. Therefore:

$$\sum_{m \in R^{k,l}} y_m = \mathbf{1}^{k,l} \cdot \mathbf{y} = \sum_i \left(\frac{\mathbf{1}^{k,l} \cdot v^i}{|\mathbf{1}^{k,l}|} \right) v^i \cdot \mathbf{y}.$$

The second equality is given by decomposition of $1^{k,l}$ into the basis of the eigenvectors v^i . On the right-hand side, the terms $1^{k,l} \cdot v^i$ equal 0 anytime v^i is associated with categories that are either disjoint with $R^{k,l}$ or its subsets, due to [equation \(20\)](#) resp. [equation \(21\)](#). The nonzero terms that remain are thus given by i 's that are associated with levels $(l + 1)$ and higher. The sum of such quantities is constant if and only if $v^i \cdot \mathbf{y}$ is constant for all such i , which it is if and only if $\lambda \geq \lambda_l$, where we use step 3 for $(l + 1)$. \square

Proof of Corollary 1.

Part (i): Since $y = H\tilde{\mathbf{x}} = \frac{\Theta^{-1}}{2}\tilde{\mathbf{x}}$, we get

$$y_m = 2 \sum_i \Lambda_i (v^i \cdot \tilde{\mathbf{x}}) v_m^i = 2 \sum_i \Lambda_i E[v^i \cdot \mathbf{x}] v_m^i.$$

Because $\lambda > \lambda_2$, information is acquired about $(v^i \cdot \mathbf{x})$ only if v^i is associated with level $l = 2$, and not with $l = 3$. Therefore, y_m can depend on x_n only via $E[v^i \cdot \mathbf{x}] v_m^i$ for such v^i associated with $l = 2$, and that is via those that are not associated with R , because $n \notin R$. For such i , due to [equation \(20\)](#), $v_m^i = 0$, and thus each quantity $E[v^i \cdot \mathbf{x}] v_m^i$ equals 0 or $E[v^i \cdot \mathbf{x}]$ does not depend on x_n . In either case, y_m does not depend on x_n .

Part (ii): This is an immediate implication of [Proposition 3](#).

Part (iii): Similarly to the above, $E[y_m | \mathbf{x}] = (HE[\sum_i (v^i \cdot \mathbf{x}) v^i | \mathbf{x}]) = 2 \sum_i \xi_i \Lambda_i (v^i \cdot \mathbf{x}) v_m^i$, where $\xi_i = \max(0, 1 - \frac{\lambda}{\Lambda_i \sigma_0^2})$, due to [equation \(7\)](#). The derivative of interest then is:

$$\begin{aligned} \frac{\partial E[y_m | x_m - x_n]}{\partial (x_m - x_n)} &= 2 \frac{\partial \sum_i \xi_i \Lambda_i (v^i \cdot \mathbf{x}) v_m^i}{\partial (x_m - x_n)} \\ &= 2 \sum_i \xi_i \Lambda_i v_m^i (v_m^i - v_n^i) = \sum_i \xi_i \Lambda_i (v_m^i - v_n^i)^2. \end{aligned}$$

The last equality is implied by symmetry of the setup, where $\frac{\partial E[y_m | x_m - x_n]}{\partial (x_m - x_n)} = \frac{\partial E[y_n | x_n - x_m]}{\partial (x_n - x_m)}$, and thus

$$\begin{aligned} 2 \sum_i \xi_i \Lambda_i v_m^i (v_m^i - v_n^i) &= \sum_i \xi_i \Lambda_i v_m^i (v_m^i - v_n^i) + \xi_i \Lambda_i v_n^i (v_n^i - v_m^i) \\ &= \sum_i \xi_i \Lambda_i (v_m^i - v_n^i)^2. \end{aligned}$$

The derivative is constant (i.e., the dependence is linear) and clearly nonnegative. Moreover, it is positive since for i 's associated with $l = 2$ the attention factor ξ_i is positive and due to [equation \(21\)](#) the nonzero eigenvectors cannot have entries v_m^i constant for all $m \in R$. \square

FACT 2. If $\lambda = 0$, then $\frac{\partial y_m}{\partial x_n} < 0$ for all $m \neq n$.

Proof of Fact 2. Because $y = H\mathbf{x} = \frac{\Theta^{-1}}{2}\mathbf{x}$, we need to show that the off-diagonal elements of $\frac{\Theta^{-1}}{2}$ are negative. We get:

$$(26) \quad \frac{\partial y_m}{\partial x_n} = \frac{\partial(H\mathbf{x})_m}{\partial x_n} = 2 \frac{\partial \sum_i \Lambda_i (v^i \cdot \mathbf{x}) v_m^i}{\partial x_n} = 2 \sum_i \Lambda_i v_n^i v_m^i,$$

where Λ_i are eigenvalues and v^i orthonormal eigenvectors of $\Omega = \frac{\Theta^{-1}}{4}$.

Let $l^* > 1$ be the smallest level such that both m, n belong to a category R^{k,l^*} . From [Lemma 1](#), we know that $v_n^i v_m^i$ on the right-hand side of [equation \(26\)](#) is equal to 0 for all i associated with a level lower than l^* , by condition (20), and $v_n^i v_m^i$ is nonnegative for all i associated with levels higher than l^* , by condition (22).

Next we note that $\sum_i v_n^i v_m^i = 0$. This is because the orthonormal eigenvectors form a unitary matrix as its rows. The columns of such a unitary matrix are then also orthonormal. Therefore,

$$-\sum_{i \sim l^*} v_n^i v_m^i = \sum_{i \sim > l^*} v_n^i v_m^i > 0,$$

where $i \sim l^*$ denotes i 's associated with the level l^* , and $i \sim > l^*$ those associated with levels higher than l^* . Let Λ^* denote the eigenvalue Λ_i associated with the level l^* . We then get:

$$(27) \quad \begin{aligned} \frac{\partial y_m}{\partial x_n} &= 2 \sum_i \Lambda_i v_n^i v_m^i = 2\Lambda^* \sum_{i \sim l^*} v_n^i v_m^i + 2 \sum_{i \sim > l^*} \Lambda_i v_n^i v_m^i \\ &= 2 \sum_{i \sim > l^*} (\Lambda_i - \Lambda^*) v_n^i v_m^i < 0. \end{aligned}$$

The inequality on the right-hand side holds because, for substitutes, Λ_i is decreasing in the level l^* , see [equation \(25\)](#), and because $v_n^i v_m^i \geq 0$ for all levels higher than l^* . \square

COROLLARY 3. The derivative $\frac{\partial E[y_m - y_n | x_m - x_n]}{\partial(x_m - x_n)}$ is decreasing in the level l^* of the smallest common category of m and n . The monotonicity is strict if $\lambda < \lambda_{l^*-1}$.

Proof of Corollary 3. First, similarly to the proof of Proposition 2:

$$E[y_m | \mathbf{x}] = \left(HE \left[\sum_i (v^i \cdot \mathbf{x}) v^i | \mathbf{x} \right] \right) = 2 \sum_i \xi_i \Lambda_i (v^i \cdot \mathbf{x}) v_m^i.$$

Now we express the derivative of interest:

$$\begin{aligned} \frac{\partial E[y_m - y_n | x_m - x_n]}{\partial(x_m - x_n)} &= 2 \frac{\partial \sum_i \xi_i \Lambda_i (v^i \cdot \mathbf{x}) (v_m^i - v_n^i)}{\partial(x_m - x_n)} \\ &= 2 \sum_i \xi_i \Lambda_i (v_m^i - v_n^i)^2, \end{aligned}$$

where $\xi_i = \max(0, 1 - \frac{\lambda}{\Lambda_i \sigma_0^2})$, due to equation (7). Note that ξ_i is weakly increasing in Λ_i .

As in the proof of Fact 2, we know that because v^i 's are orthonormal, they form rows of a unitary matrix. Columns of this unitary matrix are also orthonormal, and thus $\sum_i (v_m^i)^2 = 1$. Hence

$$2 \sum_i \xi_i \Lambda_i (v_m^i - v_n^i)^2 = const - 4 \sum_i \xi_i \Lambda_i v_m^i v_n^i.$$

Let Λ^* denote the eigenvalue Λ_i associated with the level l^* of the smallest common category of m, n , and $\xi^* = \xi_i$ for such i . Now we apply the same steps as in equation (27). Terms with products on levels lower than l^* drop out, and we get

$$-4 \sum_i \xi_i \Lambda_i v_m^i v_n^i = 4 \sum_{i \sim > l^*} (\xi^* \Lambda^* - \xi_i \Lambda_i) v_n^i v_m^i,$$

where the summation is over all i associated with levels higher than l^* . All the terms in the summation on the right-hand side are nonnegative. This is because the terms $\xi_i \Lambda_i$ are decreasing in the level for substitutes, thus $(\xi^* \Lambda^* - \xi_i \Lambda_i) \geq 0$, and because $v_n^i v_m^i$ are nonnegative for all i associated with a level higher than l^* . Increasing l^* thus decreases the sum. For a higher l^* , nonnegative terms for some i drop out completely, and the remaining terms

are scaled down because Λ^* decreases. Therefore, increasing l^* weakly decreases $\frac{\partial E[y_m - y_n | x_m - x_n]}{\partial (x_m - x_n)}$.

Moreover, if $\xi^* > 0$, then the monotonicity is strict. In that case, $(\xi^* \Lambda^* - \xi_i \Lambda_i) > 0$ and for all levels $l > l^*$ there always exists i such that $v_m^i v_n^i > 0$, because on all such levels there exists a category including both m and n . Therefore, there exists i on the level $(l^* + 1)$ such that $(\xi^* \Lambda^* - \xi_i \Lambda_i) v_m^i v_n^i$ is positive. Dropping this term as l^* increases decreases the derivative strictly. \square

Proof of Proposition 4. The variance-covariance matrix of posterior means (describing correlations of beliefs about x_i and x_j) is $P = (\psi - \Sigma)$. This matrix is diagonal in the basis of eigenvectors v^k , that is, $P = UQU^{-1}$, where the columns of U are v^i . The diagonal elements of $Q_{kk} \equiv Q_k$ equal $\sigma_0^2 - \sigma_k^2$, which is the reduction of uncertainty about $v^k \cdot \mathbf{x}$. The reduction $Q_k = \max(0, \sigma_0^2 - \frac{\lambda}{2\Lambda_k})$ is weakly increasing in Λ_k and weakly decreasing in λ , see Proposition 1.

The resulting variance-covariance matrix of actions is $A = HPH^{-1}$, where $P_{ij} = \sum_k Q_k v_i^k v_j^k$, and v^k are eigenvectors of H with eigenvalues $2\Lambda_k$. The matrix A thus is:²³

$$(28) \quad A_{ij} = \sum_k Q_k (2\Lambda_k)^2 v_i^k v_j^k.$$

Finally, the correlation of interest, that of y_m and $Y_{-m} = \sum_{n \in R^{k,l} \setminus m} y_n$, is given by

$$\begin{aligned} \frac{Cov(y_m, Y_{-m})}{\sqrt{Var(y_m)Var(Y_{-m})}} &= \frac{(n-1)A_{i,j}}{\sqrt{A_{ii}((n-1)(n-2)A_{i,j} + (n-1)A_{ii})}} \\ &= \frac{\sqrt{n-1}\rho_{ij}}{\sqrt{(n-2)\rho_{ij} + 1}}, \end{aligned}$$

where $\rho_{ij} = \frac{A_{ij}}{\sqrt{A_{ii}A_{jj}}} = \frac{A_{ij}}{A_{ii}}$ is the correlation between y_i and y_j such that $i \neq j$. The correlation of y_m and Y_{-m} is thus an increasing function of ρ_{ij} . To prove the statement of Proposition 4, it now suffices to show that ρ_{ij} is decreasing in λ .

23. Equation (28) implies that actions y_i and y_j are more positively correlated if more uncertainty is reduced in the direction of v^k , for which the signs of entries v_i^k and v_j^k are the same.

Next, using [equation \(28\)](#) we express the derivative of the correlation:

$$(29) \quad \frac{\partial \rho_{ij}}{\partial \lambda} = \frac{\left(-\sum_k 2\Lambda_k v_i^k v_j^k\right) \left(\sum_k Q_k (2\Lambda_k)^2 (v_i^k)^2\right) - \left(\sum_k Q_k (2\Lambda_k)^2 v_i^k v_j^k\right) \left(-\sum_k 2\Lambda_k (v_i^k)^2\right)}{\sum_k Q_k (2\Lambda_k)^2 (v_i^k)^2},$$

where the sums are over all k such that $\lambda < 2\Lambda_k$. Due to [Lemma 1](#), the eigenvectors can be selected such that $v_i^k v_j^k = -1$ for some vectors v^k that are associated with the smallest level on which goods i and j are in the same category, and let the level be l^* and the number of such vectors be ψ_{l^*} . Similarly, $v_i^k v_j^k = 1$ for some v^k that are associated with levels higher than l^* , and let ψ_s be the number of such vectors on the level s . For all other vectors v^k : $v_i^k v_j^k = 0$.

Let $\hat{L} \geq (l + 1)$ be the largest s such that $\lambda < 2\Lambda_s$. The numerator of the right-hand side of [equation \(29\)](#) then equals eight times the following quantity:

$$\begin{aligned} & \left(\psi_{l^*} \Lambda_{l^*} - \sum_{s=l^*+1}^{\hat{L}} \psi_s \Lambda_s \right) \left(\sum_{s=l^*}^{\hat{L}} \psi_s Q_s (\Lambda_s)^2 \right) \\ & - \left(-\psi_{l^*} Q_{l^*} (\Lambda_{l^*})^2 + \sum_{s=l^*+1}^{\hat{L}} \psi_s Q_s (\Lambda_s)^2 \right) \left(-\sum_{s=l^*}^{\hat{L}} \psi_s \Lambda_s \right) \\ & = \psi_{l^*} \Lambda_{l^*} \left(\sum_{s=l^*+1}^{\hat{L}} \psi_s Q_s (\Lambda_s)^2 \right) - \psi_{l^*} Q_{l^*} (\Lambda_{l^*})^2 \left(\sum_{s=l^*+1}^{\hat{L}} \psi_s \Lambda_s \right) \\ & = \psi_{l^*} \Lambda_{l^*} \sum_{s=l^*+1}^{\hat{L}} \left(\psi_s Q_s (\Lambda_s)^2 - \psi_s Q_{l^*} \Lambda_s \Lambda_{l^*} \right) < 0. \end{aligned}$$

In the last step we used the fact that $Q_s \Lambda_s$ are decreasing in the level s , see the proof of [Proposition 3](#). This together with the positivity of the denominator of the right-hand side of [equation \(29\)](#) concludes the proof. \square

PROPOSITION 9. Let $L = 2$ and $N = 4$, and let $\sigma_{0,m}^2$ denote the prior variance of x_m . If $\sigma_{0,1}^2 \neq \sigma_{0,2}^2 = \sigma_{0,3}^2 = \sigma_{0,4}^2$, then there exist α , $\lambda_1 < \lambda_2$ such that for $\lambda \in (\lambda_1, \lambda_2)$

$$\alpha y_1 + y_2 + y_3 + y_4 = \text{constant},$$

while individual y_m 's are not constant. Moreover, $\alpha > 1$ if and only if $\sigma_{0,1}^2 < \sigma_{0,2}^2$.

Proof of Proposition 9. Without loss of generality $\sigma_{0,2}^2 = 1$. We first transform the state space such that in the new coordinates, $\tilde{x}_1 = \frac{x_1}{\sqrt{\sigma_{0,1}^2}}$ and $\tilde{x}_m = x_m$ for all $m > 1$; the new prior variance-covariance matrix is then $\tilde{\Psi} = I$. The only other change to the original choice problem is that now the utility is $-\mathbf{y}'\Theta\mathbf{y} + \tilde{\mathbf{x}}'\tilde{B}\mathbf{y}$, where the element \tilde{B}_{11} of the new matrix \tilde{B} interacting actions and states is equal to $\alpha = \sqrt{\sigma_{0,1}^2}$.

Now we compute the loss matrix $\Omega = \frac{\tilde{B}\Theta^{-1}\tilde{B}'}{4}$, its eigenvectors and eigenvalues.²⁴ The three largest eigenvalues are associated with eigenvectors v^1 , v^2 , and v^3 , which are proportional to $(0, -1, 0, 1)$, to $(0, -1, 1, 0)$, and to $(\varrho^1, 1, 1, 1)$ respectively. The eigenvector associated with the strictly smallest eigenvalue is $v^4 \propto (\varrho^2, 1, 1, 1)$, where $\varrho^1 \neq \varrho^2$.

Proposition 1 implies that there exist $\lambda_2 > \lambda_1 > 0$ such that for $\lambda \in (\lambda_1, \lambda_2)$ the agent acquires information about $(v^i \cdot \mathbf{x})$ for $i \in \{1, 2, 3\}$ and does not acquire any information about $(v^4 \cdot \mathbf{x})$. This implies that $(v^i \cdot \tilde{\mathbf{x}})$ varies for $i \in \{1, 2, 3\}$, while $(v^4 \cdot \tilde{\mathbf{x}})$ is constant. In this case then:

$$\mathbf{y} = \tilde{H} \left(\sum_{i \in \{1, \dots, 4\}} (v^i \cdot \tilde{\mathbf{x}}) v_i \right) = \sum_{i \in \{1, 2, 3\}} (v^i \cdot \tilde{\mathbf{x}}) \tilde{H} v_i + \mathbf{y}^0,$$

where \mathbf{y}^0 is a constant vector and $\tilde{H} = \frac{\Theta^{-1}\tilde{B}'}{2}$. The consumption vector \mathbf{y} does not span all of \mathbb{R}^4 , but only its three-dimensional subspace.

Thus, we look for a vector $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ that is orthogonal to $\tilde{H}v^1$, $\tilde{H}v^2$, and $\tilde{H}v^3$, which span variations in \mathbf{y} . Straight-forward algebra reveals that $\tilde{H}v^1$ and $\tilde{H}v^2$ are proportional to $(0, -1, 0, 1)$, resp. to $(0, -1, 1, 0)$, and thus we know that a vector orthogonal to these two vectors must satisfy $\alpha_2 = \alpha_3 = \alpha_4$. Let us normalize $\alpha_2 = 1$, and we get

$$\alpha = (\alpha_1, 1, 1, 1).$$

24. We provide all the detailed computations on request.

Notice also that this vector is not orthogonal to the canonical vectors of \mathbb{R}^4 , and thus y_m are not constant. The condition of orthogonality of $(\mathbf{y} - \mathbf{y}^0)$ and α implies:

$$0 = (\mathbf{y} - \mathbf{y}^0) \cdot (\alpha_1, 1, 1, 1) = \alpha_1 y_1 + y_2 + y_3 + y_4 + Y^0,$$

where $Y^0 = -\mathbf{y}^0 \cdot (\alpha_1, 1, 1, 1)$ is a constant, and thus $\alpha_1 y_1 + y_2 + y_3 + y_4$ is constant.

Finally, we need to show that $\alpha_1 > 1$ if and only if $\alpha < 1$. We express α_1 using the condition that $\tilde{H}v^3$ is orthogonal to α , and tedious but basic algebra reveals that in fact $\alpha > 1$ if and only if $\alpha < 1$. \square

One natural question is whether the agent still engages in soft budgeting in the sense of the text, that an increase in the consumption of a good is associated with a decrease in the consumption of other goods much more than with full information. To measure this, we analyze the relative volatility of the budget, that is, ratio of the variance of $\sum_i y_i$ and the sum of variances of the y_i . Intuitively, this answers how much the agent changes her budget relative to how much she changes the consumption levels of the individual goods. For $\lambda = 0.3$, for instance, this ratio is 0 for $\alpha = 1$ and increasing fairly slowly. For $\theta = \frac{1}{4}$ we find that the volatility of x_1 needs to be quadrupled, that is, $\sigma_{0,1}^2 = 4\sigma_{0,2}^2$, to make the relative volatility of the budget one-half of the relative volatility under perfect information.

Proof of Proposition 5. The first step is the same as in the proof of Proposition 3. In the second, we first note that the ordering of the eigenvalues is the opposite since for complements: $\gamma^2 < \dots < \gamma^L < 0$. Therefore, we find that there are $\lambda_1, \dots, \lambda_L$ satisfying $\lambda_1 < \dots < \lambda_L$ such that $v^i \cdot \mathbf{x}$ is constant for all v^i associated with level l , if and only if $\lambda \geq \lambda_{l-1}$.

Finally, let us express the difference between y_m and y_n , where $m, n \in R^{k,l}$ belong to the same category k on the level l .

$$y_m - y_n = (H\tilde{\mathbf{x}})_m - (H\tilde{\mathbf{x}})_n = 2 \sum_i \Lambda_i (v^i \cdot \tilde{\mathbf{x}})(v_m^i - v_n^i),$$

where we decomposed $\tilde{\mathbf{x}}$ into the basis of $\{v^i\}$. Condition (22) states that if m, n are both in the same category on a level lower than the one i is associated with, then $(v_m^i - v_n^i) = 0$.

Because $m, n \in R^{k,l}$, then the only nonzero elements on the right-hand side might be those with i that is associated with a level

l and lower. Above we showed that $v^i \cdot \tilde{\mathbf{x}}$ associated with levels l and lower are constant if and only if $\lambda \geq \lambda_{l-1}$. Therefore, $y_m - y_n$ is constant if and only if $\lambda \geq \lambda_{l-1}$. The symmetry of the problem in fact implies $y_m = y_n$ if and only if $\lambda \geq \lambda_{l-1}$, and relabeling of the threshold costs $\lambda_{l-1} \rightarrow \lambda_l$ concludes the proof. \square

COROLLARY 4. Let $L = 3$ such that $R^{1,2} = \{1, 2\}$ and $R^{2,2} = \{3, 4\}$, where $\gamma^2 < 0$ and $\gamma^3 > 0$. Then there exist $\hat{\lambda}_2 < \hat{\lambda}_1$ such that for all $\lambda \in (\hat{\lambda}_2, \hat{\lambda}_1)$: individual y_m are not constant, but

$$(30) \quad y_1 = y_2, \quad y_3 = y_4, \quad \text{and} \quad (y_1 + y_2 + y_3 + y_4) = \text{constant}.$$

Proof of Corollary 4. Applying the arguments of the proof of Proposition 3 for the level $l = 3$ of substitutes (categories $R^{1,2}$ and $R^{2,2}$) we find that there exist $\lambda_2 > \lambda_3$ such that $y_1 + y_2 + y_3 + y_4 = \text{const}$ if and only if $\lambda \geq \lambda_3$, and $y_1 + y_2 = y_3 + y_4 = \text{const}$ if and only if $\lambda \geq \lambda_2$.

Similarly, replicating the proof of Proposition 5 for the level $l = 2$ of complements we find that there exists $\lambda_1 > 0$ such that $y_1 = y_2 = \text{const}$ and $y_3 = y_4 = \text{const}$ if and only if $\lambda \geq \lambda_1$.

Finally, from equation (24) we find that that $\mu^2 = 1 - \gamma^2 > 1$, because for complements $\gamma^2 < 0$, and $\mu^3 = \mu^2 + (\gamma^2 - \gamma^3) = 1 - \gamma^3 < 1$ because for substitutes $\gamma^3 > 0$. Therefore $\mu^2 > \mu^3$, and thus $\lambda_1 < \lambda_2$. The statement is concluded by denoting $\hat{\lambda}_1 = \max(\lambda_1, \lambda_3)$ and $\hat{\lambda}_2 = \lambda_2$. Hence for $\lambda \in (\hat{\lambda}_1, \hat{\lambda}_2)$: $y_1 + y_2 + y_3 + y_4 = \text{const}$, $y_1 = y_2 = \text{const}$ and $y_3 = y_4 = \text{const}$, while $y_1 + y_2 \neq y_3 + y_4$, and thus individual y_m are not constant. \square

LEMMA 2. *Losses from uncertainty of a fixed form:* for utility function $-\mathbf{y} \odot \mathbf{y} + (\tilde{\mathbf{x}} - \mathbf{p})' \cdot \mathbf{y}$, losses from imperfect information of a fixed form about $\mathbf{p} \cdot v^i$ are lower when thinking in terms of spending than when thinking in terms of consumption if and only if

$$\epsilon^i > \frac{1}{2}.$$

Proof of Lemma 2.

We first express the analog of expected losses from imperfect information, equation (15), for the spending choice variables using

transformation (6). Let $\mathbf{p}' = \mathbf{p} - \bar{\mathbf{x}}$.

$$\begin{aligned}
 U(\mathbf{u}, \mathbf{p}) &= -\left(\mathbf{y} + \frac{\Theta^{-1}}{2}\mathbf{p}'\right)' \Theta \left(\mathbf{y} + \frac{\Theta^{-1}}{2}\mathbf{p}'\right) + \frac{\mathbf{p}'\Theta^{-1}\mathbf{p}'}{4} \\
 &= -\frac{1}{\bar{\mathbf{p}}^2} \left(\mathbf{Y} - \bar{\mathbf{y}}(\mathbf{p} - \bar{\mathbf{p}}) + \frac{\bar{\mathbf{p}}\Theta^{-1}}{2}\mathbf{p}'\right)' \Theta \left(\mathbf{Y} - \bar{\mathbf{y}}(\mathbf{p} - \bar{\mathbf{p}}) + \frac{\bar{\mathbf{p}}\Theta^{-1}}{2}\mathbf{p}'\right) \\
 (31) \quad &+ \frac{\mathbf{p}'\Theta^{-1}\mathbf{p}'}{4}.
 \end{aligned}$$

Therefore, the optimal spending \mathbf{Y} conditional on posterior beliefs (with a mean $\bar{\mathbf{p}}^?$) is:

$$\mathbf{Y} = \bar{\mathbf{y}}(\mathbf{p} - \bar{\mathbf{p}}) - \frac{\bar{\mathbf{p}}\Theta^{-1}}{2}(\mathbf{p} - \bar{\mathbf{x}}).$$

The utility loss from imperfect beliefs is thus equal to:

$$(32) \quad (\bar{\mathbf{p}} - \mathbf{p})' \Omega^N (\bar{\mathbf{p}} - \mathbf{p}),$$

where

$$\Omega^N = \frac{1}{\bar{\mathbf{p}}^2} \left(\frac{\bar{\mathbf{p}}\Theta^{-1}}{2} - \bar{\mathbf{y}}\right)' \Theta \left(\frac{\bar{\mathbf{p}}\Theta^{-1}}{2} - \bar{\mathbf{y}}\right),$$

which can be rearranged to:

$$(33) \quad \Omega^N = \Omega - \left(\frac{\bar{\mathbf{y}}}{\bar{\mathbf{p}}}\right) I + \left(\frac{\bar{\mathbf{y}}}{\bar{\mathbf{p}}}\right)^2 \Theta.$$

The matrix $\Omega = \frac{\Theta^{-1}}{4}$, as defined right under equation (5). The loss matrices Ω and Ω^N have the same eigenvectors because Θ and Θ^{-1} have the same eigenvectors. However, their eigenvalues Λ_i and Λ_i^N , which also drive the extent of losses, can differ:

$$(34) \quad \Lambda_i^N = \Lambda_i - \left(\frac{\bar{\mathbf{y}}}{\bar{\mathbf{p}}}\right) + \left(\frac{\bar{\mathbf{y}}}{\bar{\mathbf{p}}}\right)^2 \frac{1}{4\Lambda_i} = \Lambda_i - \left(\frac{\bar{\mathbf{y}}}{\bar{\mathbf{p}}}\right) \left(1 - \frac{\bar{\mathbf{y}}}{\bar{\mathbf{p}}} \frac{1}{4\Lambda_i}\right).$$

Spending choice variables thus imply a lower eigenvalue associated with v^i , and according to equation (6) lower losses in this direction if and only if

$$(35) \quad \frac{4\bar{\mathbf{p}}\Lambda_i}{\bar{\mathbf{y}}} > 1.$$

To provide interpretation to this expression, let us introduce the elasticity of consumption with respect to the price eigenvector v^i , that is, the ratio of relative changes of consumption with respect to relative changes of prices along v^i (note that $Hv^i = \frac{\Theta^{-1}}{2} = 2\Lambda_i v^i$),

$$(36) \quad \epsilon^i = \frac{\partial(\mathbf{y} \cdot v^i | \mathbf{p} \cdot v^i = x)}{\partial x} \frac{\bar{p}}{\bar{y}} = \frac{2\bar{p}}{\bar{y}} \Lambda_i.$$

Condition (35) then takes the form of:

$$(37) \quad \epsilon^i > \frac{1}{2}. \quad \square$$

Proof of Proposition 6. Part (i) is an immediate implication of Lemma 2. This is because if both $\epsilon^1, \epsilon^2 > \frac{1}{2}$, then losses are lower when thinking in terms of spending for any given form of information. Therefore, whatever information strategy the agent chooses when thinking in terms of consumption, the agent can generate a higher objective when thinking in terms of spending by replicating the same information strategy.

Part (ii) is more involved. Consider the decomposition into one-dimensional problems as in equation (6). The objective is then

$$(38) \quad \sum_i \max \left(-\Lambda_i \sigma_0^2, -\frac{\lambda}{2} - \frac{\lambda}{2} \log \frac{2\Lambda_i \sigma_0^2}{\lambda} \right).$$

The first element in the bracket is the objective if no information is processed, while the second is the utility from imperfect posterior beliefs less the cost of information.

If $\lambda > \lambda_1$, that is, no information is processed, then the difference between the objective under spending and under consumption is $\sum_i (\Lambda_i^N - \Lambda_i) \sigma_0^2$, which equals

$$(39) \quad \sigma_0^2 \Lambda_1 \frac{2}{\epsilon^1} \left(1 - \frac{1}{2\epsilon^1} \right) (N - 1) + \sigma_0^2 \Lambda_2 \frac{2}{\epsilon^2} \left(1 - \frac{1}{2\epsilon^2} \right).$$

where we used equation (42) to express Λ_i^N in terms of ϵ^i .

If $\lambda_2 < \lambda < \lambda_1$ then under consumption the agent processes information about $\mathbf{x} \cdot v^1$, but does not process information about $\mathbf{x} \cdot v^2$. We now express the difference between the objective under spending and under consumption when in both cases the information acquisition is optimal for the problem with consumption. The

difference is:

$$(40) \quad \frac{\lambda}{2} \frac{2}{\epsilon^1} \left(1 - \frac{1}{2\epsilon^1}\right) (N-1) + \sigma_0^2 \Lambda_2 \frac{2}{\epsilon^2} \left(1 - \frac{1}{2\epsilon^2}\right).$$

The difference between objectives when information is chosen optimally under spending, too, is thus at least as high as this quantity. The second term in [equation \(40\)](#) is the same as in [equation \(39\)](#) since no information is processed about $\mathbf{x} \cdot v^2$ in either case. However, the first term is $-\frac{\lambda}{2} - \frac{\lambda}{2} \log \frac{2\Lambda_1\sigma_0^2}{\lambda}$ for consumption and $-\frac{\lambda}{2} \frac{\Lambda_1^N}{\Lambda_1} - \frac{\lambda}{2} \log \frac{2\Lambda_i\sigma_0^2}{\lambda}$ for spending. The cost of information is the same in both cases, and drops out, and the losses from the same posterior beliefs are scaled by the corresponding eigenvalues.

Finally, if $\lambda < \lambda_2$, then the difference between the objectives under spending and consumption is higher than

$$(41) \quad \frac{\lambda}{2} \frac{2}{\epsilon^1} \left(1 - \frac{1}{2\epsilon^1}\right) (N-1) + \frac{\lambda}{2} \frac{2}{\epsilon^2} \left(1 - \frac{1}{2\epsilon^2}\right),$$

which is again the difference between the objectives for information under spending being held at the optimal information under consumption.

All three differences between the two objectives in [expressions \(39\)–\(41\)](#) are for $\epsilon^1 > \frac{1}{2}$ positive for sufficiently large N . In each of the expressions, the second term is independent of N , while the first terms are positive for $\epsilon^1 > \frac{1}{2}$ and increasing linearly with N . \square

Proof of Proposition 7. We replicate the proof of [Proposition 3](#) as long as the ordering of eigenvalues is the same regardless of whether thinking in terms of spending or consumption.

Plugging [equation \(36\)](#) into [equation \(34\)](#) we get

$$(42) \quad \Lambda_i^N = \Lambda_i \left(1 - \frac{2}{\epsilon^i} \left(1 - \frac{1}{2\epsilon^i}\right)\right).$$

Using [equation \(34\)](#) we can express differences between eigenvalues for nominal variables:

$$\Lambda_i^N - \Lambda_j^N = (\Lambda_i - \Lambda_j) \left(1 - \frac{\left(\frac{\bar{y}}{2\bar{p}}\right)^2}{\Lambda_i \Lambda_j}\right).$$

The right-hand side has the same sign as $(\Lambda_i - \Lambda_j)$, that is, the ordering is the same for both decision variables, if and only if

$$\epsilon^i \epsilon^j > 1.$$

If this condition holds, then the analog of Proposition 3 applies. \square

Proof of Corollary 2. We normalize $\bar{p} = 1$. Substituting our approximation $y_m = Y_m - (p_m - 1)\bar{y}$ into the agent's utility function, dropping terms the agent cannot influence, and rearranging gives the objective function

$$(43) \quad - \sum_m Y_m^2 - \sum_{m \neq n} \theta Y_m Y_n + \underbrace{\sum_m \left(\bar{x} - 1 + 2(p_m - 1)\bar{y} + \sum_{n \neq m} 2(p_n - 1)\bar{y} \right) Y_m}_{\equiv X_m}.$$

Denote by \tilde{X}_m the agent's posterior mean of X_m , and let $\mathbf{Y} = (Y_1, \dots, Y_N)'$, $\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_N)'$. We know that $\mathbf{Y} = \frac{\Theta^{-1}\tilde{\mathbf{X}}}{2}$, so $E[\mathbf{Y}] = \frac{\Theta^{-1}E[\tilde{\mathbf{X}}]}{2}$. Notice that $X_m - X_n = 2(1 - \theta)(p_m - p_n)$. Since $\lambda_2 \leq \lambda < \lambda_1$, the agent acquires information about $p_m - p_n$, which is equivalent to acquiring information about $X_m - X_n$ but not about the sum of the X_m . Hence, a decrease in p_m lowers $E[\tilde{X}_m]$ and increases $E[\tilde{X}_n]$ for all $n \neq m$, leaving the sum unchanged. This lowers $E[Y_m]$ and raises $E[Y_n]$ for all $n \neq m$. \square

Proof of Proposition 8. The expected utility from choosing consumption y_m is

$$(44) \quad \begin{aligned} & y_m - E[s(p(y_m) + x_m) - (1 - s)p(y_m + x_m)] \\ & = y_m - sp(y_m) - (1 - s)E[p(y_m + x_m)]. \end{aligned}$$

Because $p(\cdot)$ is strictly convex, $E[p(y_m + x_m)] > p(y_m)$. Hence, the expected utility from choosing y_m is strictly increasing in s . As a result, the maximum of the above expression is also strictly increasing in s .

Let $q(\cdot)$ be the inverse of $p(\cdot)$. Note that $q(\cdot)$ is strictly concave. The expected utility from choosing spending Y_m is

$$(45) \quad \begin{aligned} & E[sq(Y_m - x_m) + (1 - s)(q(Y_m) - x_m)] - Y_m \\ & = sE[q(Y_m - x_m)] + (1 - s)q(Y_m) - Y_m. \end{aligned}$$

Because $q(\cdot)$ is strictly concave, $E[q(Y_m - x_m)] < q(Y_m)$. Hence, the expected utility from choosing Y_m is strictly decreasing in s . As a result, the maximum of the above expression is also strictly decreasing in s .

To complete the proof, we show that for $s = 0$ choosing spending is optimal, and for $s = 1$ choosing quality is optimal. If $s = 0$, then expression (44) is strictly less than $y_m - p(y_m)$, which is exactly expression (45) with $Y_m = p(y_m)$, so fixing spending dominates fixing the quality. If $s = 1$, then expression (45) is strictly less than $q(Y_m) - Y_m$, which is exactly expression (44) for $y_m = q(Y_m)$, so fixing the quality dominates fixing spending. \square

CENTRAL EUROPEAN UNIVERSITY

CERGE-EI, A JOINT WORKPLACE OF CHARLES UNIVERSITY IN PRAGUE
AND THE ECONOMICS INSTITUTE OF THE CZECH ACADEMY OF SCIENCES

REFERENCES

- Abeler, Johannes, and Felix Marklein, "Fungibility, Labels, and Consumption," *Journal of the European Economic Association*, 15 (2016), 99–127.
- Afrouzi, Hassan, and Choongryul Yang, "Dynamic Rational Inattention and the Phillips Curve," Mimeo, Columbia University, 2019.
- Ameriks, John, Andrew Caplin, and John Leahy, "Wealth Accumulation and the Propensity to Plan," *Quarterly Journal of Economics*, 118 (2003), 1007–1047.
- Antonides, Gerrit, I. Manon De Groot, and W. Fred Van Raaij, "Mental Budgeting and the Management of Household Finance," *Journal of Economic Psychology*, 32 (2011), 546–555.
- Benartzi, Shlomo, and Richard H. Thaler, "Naive Diversification Strategies in Defined Contribution Saving Plans," *American Economic Review*, 91 (2001), 79–98.
- , "Heuristics and Biases in Retirement Savings Behavior," *Journal of Economic Perspectives*, 21 (2007), 81–104.
- Caplin, Andrew, and Mark Dean, "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 105 (2015), 2183–2203.
- Cheema, Amar, and Dilip Soman, "Malleable Mental Accounting: The Effect of Flexibility on the Justification of Attractive Spending and Consumption Decisions," *Journal of Consumer Psychology*, 16 (2006), 33–44.
- Cover, Thomas M., and Joy A. Thomas, *Elements of Information Theory* (Hoboken, NJ: Wiley, 2006).
- Dean, Mark, and Nate Leigh Neligh, "Experimental Tests of Rational Inattention," Columbia University Working Paper, 2017.
- Dupas, Pascaline, "Getting Essential Health Products to their End Users: Subsidize, but How Much?," *Science*, 345 (2014), 1279–1281.
- Fosgerau, Mogens, Emerson Melo, and Matthew Shum, "Discrete Choice and Rational Inattention: A General Equivalence Result," University of Copenhagen Working Paper, 2017.
- Fulton, Chad, "Mechanics of Linear Quadratic Gaussian Rational Inattention Tracking Problems," FEDS Working Paper, 2017.
- Gabaix, Xavier, "A Sparsity-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 129 (2014), 1661–1710.

- Gabaix, Xavier, and David Laibson, "Myopia and Discounting," NBER Working Paper no. 23254, 2017.
- Galperti, Simone, "A Theory of Personal Budgeting," *Theoretical Economics*, 14 (2019), 173–210.
- Gorman, W. M., "Separable Utility and Aggregation," *Econometrica*, 27 (1959), 469–481.
- Hastings, Justine S., and Jesse M. Shapiro, "Fungibility and Consumer Choice: Evidence from Commodity Price Shocks," *Quarterly Journal of Economics*, 128 (2013), 1449–1498.
- , "How Are SNAP Benefits Spent? Evidence from a Retail Panel," *American Economic Review*, 108 (2018), 3493–3540.
- Heath, Chip, and Jack B. Soll, "Mental Budgeting and Consumer Decisions," *Journal of Consumer Research*, 23 (1996), 40–52.
- Heilman, Carrie M., Kent Nakamoto, and Ambar G. Rao, "Pleasant Surprises: Consumer Response to Unexpected In-Store Coupons," *Journal of Marketing Research*, 39 (2002), 242–252.
- Henderson, Pamela W., and Robert A. Peterson, "Mental Accounting and Categorization," *Organizational Behavior and Human Decision Processes*, 51 (1992), 92–117.
- Hines, James, and Richard Thaler, "The Flypaper Effect," *Journal of Economic Perspectives*, 9 (1995), 217–226.
- Hsiaw, Alice, "Goal Bracketing and Self-Control," *Games and Economic Behavior*, 111 (2018), 100–121.
- Huberman, Gur, and Wei Jiang, "Offering versus Choice in 401(k) Plans: Equity Exposure and Number of Funds," *Journal of Finance*, 61 (2006), 763–801.
- Kahneman, Daniel, and Amos Tversky, "Choices, Values, and Frames," *American Psychologist*, 39 (1984), 341–350.
- Koch, Alexander K., and Julia Nafziger, "Goals and Bracketing under Mental Accounting," *Journal of Economic Theory*, 162 (2016), 305–351.
- Krishnamurthy, Parthasarathy, and Sonja Prokopec, "Resisting that Triple-Chocolate Cake: Mental Budgets and Self-Control," *Journal of Consumer Research*, 37 (2010), 68–79.
- Lave, Jean, *Cognition in Practice: Mind, Mathematics, and Culture in Everyday Life* (Cambridge: Cambridge University Press, 1995).
- Lin, Judy T., Christopher Bumcrot, Tippy Ulicny, Annamaria Lusardi, Gary Motola, Christine Kiefer, and Gerri Walsh, "Financial Capability in the United States 2016," Technical Report, Global Financial Literacy Excellence Center, 2016.
- Luo, Yulei, and Eric R. Young, "Signal Extraction and Rational Inattention," *Economic Inquiry*, 52 (2014), 811–829.
- Mackowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, "Rational Inattention: A Disciplined Behavioral Model," CERP Working Paper, 2018.
- Mackowiak, Bartosz, and Mirko Wiederholt, "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 99 (2009), 769–803.
- Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao, "Poverty Impedes Cognitive Function," *Science*, 341 (2013), 976–980.
- Matějka, Filip, "Rationally Inattentive Seller: Sales and Discrete Pricing," *Review of Economic Studies*, 83 (2016), 1125–1155.
- Matějka, Filip, and Alisdair McKay, "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 105 (2015), 272–298.
- Miao, Jianjun, Jieran Wu, and Eric Young, "Multivariate Rational Inattention," Boston University Working Paper, 2019.
- Morris, Stephen, and Philipp Strack, "The Wald Problem and the Equivalence of Sequential Sampling and Static Information Costs," Mimeo, MIT, 2017.
- Morris, Stephen, and Ming Yang, "Coordination and Continuous Stochastic Choice," Mimeo, MIT, 2016.

- Pagel, Michaela, "Expectations-Based Reference-Dependent Life-Cycle Consumption," *Review of Economic Studies*, 84 (2017), 885–934.
- Pomatto, Luciano, Philipp Strack, and Omer Tamuz, "The Cost of Information," Mimeo, Caltech, 2019.
- Rainwater, Lee, Richard P. Coleman, Gerald Handel, and W. Lloyd Warner, *Workingman's Wife: Her Personality, World and Life Style* (Dobbs Ferry, NY: Oceana Publications, 1959).
- Rajagopal, Priyali, and Jong-Youn Rha, "The Mental Accounting of Time," *Journal of Economic Psychology*, 30 (2009), 772–781.
- Reis, Ricardo, "Inattentive Consumers," *Journal of Monetary Economics*, 53 (2006), 1761–1800.
- Schilbach, Frank, Heather Schofield, and Sendhil Mullainathan, "The Psychological Lives of the Poor," *American Economic Review*, 106 (2016), 435–440.
- Shefrin, Hersh M., and Richard H. Thaler, "The Behavioral Life-Cycle Hypothesis," *Economic Inquiry*, 26 (1988), 609–643.
- Simonson, Itamar, "The Effect of Purchase Quantity and Timing on Variety-Seeking Behavior," *Journal of Marketing Research*, 27 (1990), 150–162.
- Sims, Christopher A., "Implications of Rational Inattention," *Journal of Monetary Economics*, 50 (2003), 665–690.
- Telatar, Emre, "Capacity of Multi-antenna Gaussian Channels," *European Transactions on Telecommunications*, 10 (1999), 585–595.
- Thaler, Richard H., "Mental Accounting and Consumer Choice," *Marketing Science*, 4 (1985), 199–214.
- , "Mental Accounting Matters," *Journal of Behavioral Decision Making*, 12 (1999), 183–206.
- Tversky, Amos, and Daniel Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science*, 211 (1981), 453–458.
- Veldkamp, Laura L., "Information Markets and the Comovement of Asset Prices," *Review of Economic Studies*, 73 (2006), 823–845.
- Verstyuk, Sergiy, "Log, Stock and Two Simple Lotteries," Mimeo, Harvard, 2019.
- Volpp, Kevin G., Leslie K. John, Andrea B. Troxel, Laurie Norton, Jennifer Fassbender, and George Loewenstein, "Financial Incentive-Based Approaches for Weight Loss," *Journal of the American Medical Association*, 300 (2008), 2631–2637.
- Woodford, Michael, "Information-Constrained State-Dependent Pricing," *Journal of Monetary Economics*, 56 (2009), S100–S124.
- , "Inattentive Valuation and Reference-Dependent Choice," Columbia University Working Paper, 2012.
- Zhang, C. Yiwei, and Abigail B. Sussman, "Perspectives on Mental Accounting: An Exploration of Budgeting and Investing," *Financial Planning Review*, 1 (2018).