

A Sparsity-Based Model of Bounded Rationality*

Xavier Gabaix

NYU Stern, CEPR and NBER

October 13, 2010

Abstract

This paper proposes and analyzes a model with boundedly rational features in which the decision-maker (DM) behaves like an economist who builds a simplified representation of the world. Crucially, this representation is “sparse,” i.e. uses few parameters that are non-zero, or differ from the usual state of affairs. The DM may imperfectly maximize, based again on a penalty related to sparsity. The lack of sparsity is formulated so as to lead to well-behaved, convex maximization problems. The model is a tractable algorithm that can be applied with paper and pencil in many situations of interest. I apply it to a variety of prototypical economic situations: hitting a target with selective attention; picking a consumption bundle, but with imperfect understanding of price; optimal pricing with boundedly rational consumers – which, when paired with optimal response by firms, generates a novel mechanism for price rigidity; life-cycle consumption and investment problems; failures of Euler equations; portfolio choice problems with stocks and flows; the “acquiring a company” problem; dollar auction game. I conclude that the model may be a useful proposal for tractable analysis of bounded rationality in economic situations.

Preliminary and Incomplete

*xgabaix@stern.nyu.edu. I thank David Laibson for a great many enlightening conversations about bounded rationality over the years. I thank Farzad Saidi for very good research assistance, and seminar participants at Cornell, NYU, Yale and the Stanford Institute for Theoretical Economics for helpful comments.

1 Introduction

This paper proposes a tractable model with boundedly rational (BR) features. It is designed to be easy to apply in concrete economic situations.

Its principles are the following. First, the decision-maker (DM) in the model is not the rational agent model, but is best thought of as an economist building a model of the world (a model-in-model or MIM). He builds representations of the world that are simple enough, and thinks about the world through his partial model. Second, and most crucially, this representation is “sparse,” i.e. uses few parameters that are non-zero, or differ from the usual state of affairs.¹ I draw from the fairly recent literature on statistics and image processing to use a notion of “sparsity” that still leads to well-behaved, convex maximization problems. Third, maximization can itself be imperfect, with a penalty that also increases as the action taken becomes too different from the default action, and it relies on the same sparsity criterion.

The DM, like an economist, simplifies his model of the world. For instance, he assumes that some parameters are just irrelevant (when they can, strictly speaking, matter a bit), and that some variables are deterministic rather than random. He assumes convenient distributions rather than the messiness of reality: e.g., he might assume a distribution with two outcomes rather than a continuum of outcomes. He models that variables are uncorrelated when they are not exactly so. These choices are controlled by an optimization of his representation of the world.

To motivate the model, I first consider the “quadratic target” problem: the DM wishes to target the sum of many variables, but does not wish to think about all of them. By Taylor expansion, this is a prototypical toy model for many optimization problems. I study how to state the cost of enriching the representation. Following antecedents in statistics and applied mathematics (Tibshirani 1996, Candès, and Tao 2006, Donoho 2006), I show that one is particularly appealing: the ℓ_1 norm, i.e. the sum of absolute values of the non-zero updates in the variables. Why? First, a quadratic cost would not generate sparsity: small updates would have a miniscule penalty, hence under that model the DM would have non-sparse representations. Second, a fixed cost per variable would give sparsity, but lose tractability; fixed costs lead to non-convex problems that make the solution very complicated in general. Instead, the ℓ_1 penalty both gives sparsity, and maintains tractability. Hence, in this quadratic loss problem, it is useful to have the penalty for lack of sparsity be the ℓ_1 norm. The model generates inattention to many variables, and dampened attention to some, as well as sparsity, so that we are on the right track.

¹The meaning of “sparse” is that of a sparse matrix or vector. For instance, a vector in $\theta \in \mathbb{R}^{100,000}$ with only a few non-zero elements is sparse.

The unweighted ℓ_1 criterion, used in the basic quadratic target problem, cannot work in the general case: for instance, dimensions might not be comparable – e.g., the units could be different. I study how to generalize it. It turns out that, under some reasonable conditions, there is only one unique algorithm that (i) penalizes the sum of absolute values in the symmetrical quadratic target problem and (ii) is invariant by changes in units and various rotations of the problem. This is the algorithm I state as the “Sparse BR” algorithm. Hence, basic invariance considerations lead to an algorithm that is fairly tightly constrained .

In addition, the algorithm features just a fairly simple optimization problem, so it is easy to apply.

I apply the model to a variety of situations. I study consumption and investment problems. The agent will pay reduced or no attention to a great many variables. This generates systematic deviations from Euler equations: they point towards inertia, as agents will react in a dampened way to many future variables. DMs react more to news about the present than to late news about the future. The marginal propensity to consume out of current income is higher than out of future income. This is much like Thaler’s “mental accounts.” The covariance of consumption with most things will be low, because of the dampening due to lack of attention, so that measured elasticity of substitutions will be low and measured risk aversion will be high.

I consider an agent buying a bundle of n goods, with an imperfect attention to all prices. The model generates a zone of insensitivity to prices: when prices are close to the default, the DM does not pay attention to them. I then study how a firm will optimally price goods sold to such BR consumers. It is clear that it will not choose just any price strictly inside the zone of consumer inattention: it will rather pick a price at its upper bound. Hence, a whole zone of prices will not be picked by firms. Even as the marginal cost of goods changes, there will be a zone of complete price rigidity. In addition, there is an asymmetry: there will be discrete jumps downwards of the price sometimes (“sales”), but not corresponding jumps up from the normal price. Hence, we get a tractable, BR-based mechanism for price rigidity.

I also show how the model gives rise to first-order risk aversion and the endowment effect. The size of the effect depends on how uncertain the good is. This is a difference from prospect theory, where the size of the effect depends only on the hedonic value of the good. Hence, the model explains why more experienced traders (List 2003) exhibit a much weaker endowment effect.

I explain how the model accounts for a series of notions about bounded rationality. For instance, the model generates epiphanies, i.e. sudden realizations of a possible state of the world. It also generates a greater tendency to adjust one’s portfolio by adjusting flows rather than stock allocations: i.e., an investor concerned that equity might be overvalued will stop buying new shares altogether (or buy fewer new shares with fresh cash), rather than sell equities

in his existing portfolio. Indeed, this suggests that empirically the difference between flow vs. stock allocation is particularly diagnostic of the outlook of BR investors.

Besides studying those economic models, or components of models, I also apply the model to some canonical laboratory games. I point out that the model is useful to interpret the experimental evidence: the “acquiring-a-company” game, the centipede game, the “buy a dollar” game.

Near the end of the paper, I describe a series of potential enrichments of the model. One is a model with constraints. I also discuss what to do when the underlying spaces are discrete rather than continuous, or when there are “domination” patterns that might (or not) be detected by the DM. These are extra tools that the DM might use in some cases. Still, a lot of the basic economics can be studied with the most basic framework.

Another extension is to multi-agent problems. To a large extent, I glue the basic Sparse-BR model to the existing ideas from the k -levels of reasoning literature (Stahl 1998, Camerer, Ho, and Chong 2004, Crawford and Iriberri 2007).

This paper tries to strike a balance between psychological realism and model tractability. The goal for the model is to be applicable without too much trouble, and at the same time to capture some dimensions of bounded rationality.

The central elements of this paper – the use of the ℓ_1 norm to model bounded rationality, the accent on sparsity, and the sparse BR algorithm – are, to the best of my knowledge, novel. I defer the discussion between this paper and rest of the literature to later in the paper, as it is best discussed when the reader knows the key elements of the model.

The plan of the paper is as follows. Section 2 motivates of the model, in the context of a stylized model where the goal is to hit a target. Section 3 states the main model. Section 4 applies the main model to a few applied problems. One is how a BR consumer picks a bundle of n goods, but doesn’t completely process the vector of prices. I also work out how a monopolist optimally sets prices given such a consumer: we will get a novel source of real price rigidity, along with occasional “sales” with large temporary changes in prices. It also shows how the model generates an endowment effect. Section 5 shows how the idea of different representations applying to simplifying random variables and categorization, using the language of “dictionaries” from the applied mathematics literature. Section 6 extends the model to multiple players. Section 7 indicates various enrichments of the model and makes the link with existing themes in behavioral economics. Section 8 discusses the limitations of this approach and concludes.

2 A Motivation: Sparsity and ℓ_1 Norm

We are developing a model where agents have sparse representations of the world, i.e. many parameters are set at “0”, the default values. To fix ideas, consider the following decision problem.

Problem 1 (*Choice Problem with Quadratic Loss*). *The random variables x_i and weights q_i are freely available, though perhaps hard to process. The problem is: Pick a to maximize*

$$u(a, q, x) = \frac{-1}{2} \left(a - \sum_{i=1}^n q_i x_i \right)^2 \quad (1)$$

If x_i 's are known, the optimal action is $a(q) = \sum_{i=1}^n q_i x_i$. However, we want to model an agent that cannot think about all these dimensions. He will just think about “the most important ones.” Hence, he will think about $a(\theta) = \sum \theta_i x_i$ for some vector θ , that endogenously has lots of zeros, i.e. is “sparse.” The expected loss is: $L = \mathbb{E}[u(a(q)) - u(a(\theta))]$, i.e., assuming for simplicity that the x_i are i.i.d. with mean 0 and variance 1:

$$L = \frac{1}{2} \sum_{i=1}^n (\theta_i - q_i)^2.$$

We will set the choice of θ as an optimization problem:

$$\min_{\theta \in \mathbb{R}^n} \frac{1}{2} \sum_i (\theta_i - q_i)^2 + \kappa \sum_i |\theta_i|^\alpha \quad (2)$$

One natural choice would be $\alpha = 2$. Then, we obtain $-(\theta_i - q_i) - 2\kappa\theta_i = 0$ i.e.

$$\theta_i = \frac{q_i}{1 + 2\kappa} \quad (3)$$

We do not get any sparsity: all features matter, with a small or large q_i . We just get some uniform dampening. Hence, we seek something else.

Another natural modelling choice is $\alpha = 0$, with the convention $|\theta|^\alpha = 1_{\theta \neq 0}$, i.e. there's a cost κ for each non-zero element. Then, the solution is:

$$\theta_i = \begin{cases} q_i & \text{if } |q_i| > \sqrt{2\kappa} \\ 0 & \text{if } |q_i| \leq \sqrt{2\kappa} \end{cases}.$$

We do obtain sparsity. However, there's a big cost in terms of tractability. Problem (2) is not convex any more when $\alpha = 0$ (it is convex if and only if $\alpha \geq 1$). Its general formulation

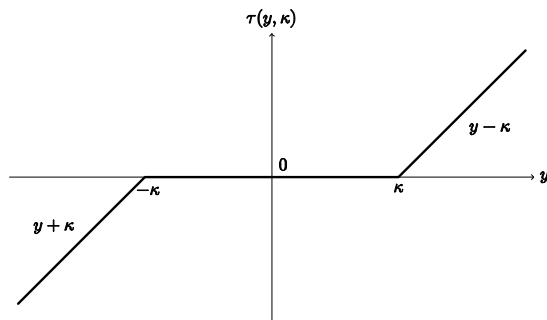


Figure 1: The τ function.

($\min_{\theta \in \mathbb{R}^n} F(\theta) + \kappa \sum_i 1_{\theta_i} \neq 0$, for a convex F) it is very hard to solve – it is NP-complete in the terminology of complexity theory (the naive solution would be to study the 2^{n_θ} subsets of non-zero elements of θ). It requires studying cases, so it considerably hampers tractability.

This leads to consider the problem with $\alpha = 1$, as argued in the recent signal-processing literature (Tibshirani 1996, Donoho 2006, Candès, and Tao 2006). Then problem (2) is convex. Let us solve it. Differentiating (2), we have:

$$-(\theta_i - q_i) - \kappa \cdot \text{sign}(\theta) = 0 \quad (4)$$

Let us solve (4) when $q_i > 0$. When the solution is $\theta_i > 0$, we obtain $\theta_i = q_i - \kappa$, which requires $q_i > \kappa$. When $0 \leq q_i \leq \kappa$, $\theta_i = 0$. In general we have:

$$\theta_i = \tau(q_i, \kappa)$$

for the truncation or “soft thresholding” function τ defined as follows and plotted in Figure 1.

Definition 1 *The truncation function τ is*

$$\tau(y, \kappa) = (|y| - \kappa)_+ \text{sign}(y) \quad (5)$$

i.e.

$$\tau(y, \kappa) = \begin{cases} y - \kappa & \text{if } y \geq \kappa \\ y + \kappa & \text{if } y \leq -\kappa \\ 0 & |y| < \kappa \end{cases} \quad (6)$$

We summarize the situation in the following Lemma.

Lemma 1 *For $A > 0$ and $K \geq 0$, the solution of*

$$\min_{\theta} \frac{A}{2} (\theta - q)^2 + K |\theta - \theta^d|$$

is

$$\theta = \theta^d + \tau \left(q - \theta^d, \frac{K}{A} \right)$$

where τ is the truncation function given in (5).

Proof. By shifting $\theta \rightarrow \theta - \theta^d$, $q \rightarrow q - \theta^d$, it is enough to consider the case $\theta^d = 0$. The f.o.c. is

$$\theta - q + \frac{K}{A} \text{sign}(\theta) = 0$$

That is, $\theta = \tau(q, K/A)$. ■

Hence, we do have some sparsity: all terms that have $|q_i| < \kappa$ are replaced by $\theta_i = 0$. For $q_i > \kappa$, we get $\theta_i = q_i - \kappa$, so there's a bit of dampening.²

The conclusion is that we can use the ℓ_1 norm, i.e. the one that corresponds to $\alpha = 1$ in (2), to generate sparsity.

I next generalize the model this sort of idea to more general functions.

3 The Basic Model

To clarify the ideas and the exposition, I start with problems with just one DM.

3.1 Model Statement

There is an action $a \in \mathbb{R}^{n_a}$, a representation $\theta \in \mathbb{R}^{n_\theta}$, and a state of the world $x \in \mathbb{R}^{n_x}$, and noise realized later $\varepsilon \in \mathbb{R}^{n_\varepsilon}$, and a value function $W(a, q, x, \varepsilon)$, $W : \mathbb{R}^{n_a} \times \mathbb{R}^{n_q} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_\varepsilon} \rightarrow \mathbb{R}$. The state of world is distributed with a known probability \mathbb{P}^x and, that given, the noise is distributed with probability $\mathbb{P}^{\varepsilon|x}$.

Suppose the DM wishes to maximize $\max_a \mathbb{E}[W(a, q, x, \varepsilon)]$, where the expectation is over the realizations of ε . The true value of the parameter is q , but people will build their model with a simpler representation θ . There is a default $a^d \in \mathbb{R}^{n_a}$ and $\theta^d \in \mathbb{R}^{n_d}$.

There is a prior knowledge of the normal variations in the action, represented by a random variable η_a , and in the representation, represented by η_θ , discussed below.

²Also, it is easy to see that, θ has no more than $\|q\|_1 / \kappa$ non-zero components.

We will use the operators on a function $f(a, \theta)$:

$$(\Delta_{\theta_i} f)(\theta) = (\theta_i - \theta_i^d) \partial_{\theta_i} f(\theta) \quad (7)$$

$$(\Delta_{a_i} f)(\theta) = (a_i - a_i^d) \partial_{a_i} f(\theta) \quad (8)$$

$$(\Delta_{\eta_\theta} f)(a) = \eta_\theta \cdot \partial_\theta f(\theta) \quad (9)$$

$$(\Delta_{\eta_a} f)(a) = \eta_a \cdot \partial_a f(a) \quad (10)$$

The notation $\partial_a f(a)$ is the differential of f at point a and the dot \cdot is the vector product; for instance, $\eta_a \cdot \partial_a f(a) = \sum_i \eta_{a_i} \frac{\partial f}{\partial a_i}(a)$.

For X a random variable, I define:

$$\|X\|_\alpha = \mathbb{E}[|X|^\alpha]^{1/\alpha} \quad (11)$$

for $\alpha \geq 0$, with the convention that $\|X\|_0 = 1_{X \neq 0}$ and $\|X\|_\infty = \text{esssup } |X|$. It is a norm when $\alpha \geq 1$, as shown by Minkowski's inequality. We will consider expressions such as $\sum_i |\theta_i| \|X_i\|_\alpha = \sum_i \|\theta_i X_i\|_\alpha$ as they combine the important ℓ_1 feature that generates tractable sparsity (the $|\theta_i|^1$ terms), and the convenience of the general $\|X_i\|_\alpha$ norm, in particular with $\alpha = 2$.

This paper proposes that the following algorithm is a useful model of agents' behavior. It may be called the "Sparse Boundedly Rational" algorithm, or "Sparse BR" for short.

Algorithm 1 (*Sparse BR Algorithm*) *To solve the problem $\max_a \mathbb{E}^x [W(a, q, x, \varepsilon)]$, the agent uses the following two steps:*

1. *Optimize on the representation of the world. Using the realism loss matrix Λ :*

$$\Lambda = -\mathbb{E}^{x, \varepsilon} [W_{a\theta} W_{aa}^{-1} W_{a\theta}] \quad (12)$$

evaluated at (a^d, θ^d) , averaged over the realizations of x and ε , determine the parametrization θ of the "model-in-model" or "representation" used by the agent as the solution of:

$$\max_\theta -\frac{1}{2} (\theta - q)' \Lambda (\theta - q) - \kappa[\theta]. \quad (13)$$

The first part is a measure of expected losses from a poor simulation, while the second part is the complexity cost of the representation,

$$\kappa[\theta] = \kappa^\theta \sum_i \|\Delta_{\theta_i} \Delta_{\eta_a} W(a, \theta, x, \varepsilon)\|_\alpha \quad (14)$$

2. **Optimize on the action.** Maximize over the action a :

$$\max_a \mathbb{E}^\varepsilon [W(a, \theta, x, \varepsilon)] - \kappa[a], \quad (15)$$

where the expectation is over the realizations of ε , and where the complexity cost of the action, $\kappa[a]$, is:

$$\kappa[a] = \kappa^a \sum_i \|\Delta_{a_i} \Delta_{\eta_\theta} W(a, \theta, x, \varepsilon)\|_\alpha. \quad (16)$$

Let me comment on the parts of the model.

First-Pass Intuition for the model When $\kappa^\theta = 0$, the DM's model of the world is the correct one: $\theta = q$. When $\kappa^a = 0$, the maximization is perfect, conditional on the model-in-model.

For many applications, it might be enough to just turn on either step 1 or step 2 of the model. In most of this paper, step 1 only will be turned on; i.e., I will assume perfect maximization given the representation of the world ($\kappa^a = 0$).

When cognition costs κ are non-zero, the model exhibits inertia and conservatism: the model-in-model (MIM) is equal to the default, and the action is equal to the default, when κ is very large. For smaller κ , the model is neither at the default nor at the costless optimal, but typically in between.

Units and scaling Sparsity penalties κ^θ and κ^a are non-dimensional. The model has the right units: equations (12)–(16) all have the dimensions of W . Also, the equations are independent of the units in which the components of θ and a are measured. However, the model is not invariant to the representations of the world θ : some will be better for the agent than others. That is probably a desirable feature of the model.

Values of η_θ and η_a Variables η_a and η_θ in part ensure that the model has the right units and scaling properties. They are typically not crucial in applications. When they're one-dimensional, we can have $\eta_a = \sigma_a$. To close the model, we can say that η_θ simply follow the distribution of q , and a follow the distribution of $a^d(q, x)$, for instance.

Why is the model set this way? The algorithm is written, first of all, to have some descriptive realism. That will be argued in the rest of the paper. Also, it was designed to be:

(i) equivalent to the discussion done in Section 2 for the quadratic problem; this is because smooth problems are locally equivalent to the quadratic target problem. The next subsection indeed shows that equivalence

(ii) dependent on $\partial_a W$, but not W directly: as the DM seeks a , the algorithm should arguably return the same answer whether we maximize $W(a, \theta, x, \varepsilon)$ and $W(a, \theta, \varepsilon) + G(\theta, x, \varepsilon)$ for an arbitrary function G . The use of $\partial_a W$ (or, in slightly disguised form, Δ_{η_a} for the non-differential case) cancels the function G throughout the model

(iii) still, dependent only on no derivative higher than the second derivatives. That's to keep the model simple and in some sense independent (at least locally) of various details like the third derivatives

(iv) invariant to the units of the components θ and a .

The Proposition, proven in the Appendix, says that there is a unique algorithm that satisfies the above four criteria. In that sense, the model is tightly constrained. The equation (14) is rather necessary.

Proposition 1 *Suppose that the determination of θ is*

$$\max_{\theta} -\frac{1}{2} (\theta - q)' \Lambda (\theta - q) - K(\theta_i - \theta_i^d, \eta_a, W_{a\theta_i}, W_{aa}, W) \quad (17)$$

for a penalty function K evaluated at the values of W and its derivatives at point (a^d, θ^d) . Suppose also that K satisfies:

(i) The value of K is unchanged under linear reparametrizations of θ_i (for $i = 1 \dots n_\theta$) and of a : for all $\lambda_i \in \mathbb{R}$, and $A \in \mathbb{R}^{n_a \times n_a}$,

$$K(\lambda_i \theta_i, A' \eta_a, W_{a\theta_i}, W_{aa}, W_a, W) = K(\theta_i, \eta_a, \lambda_i A W_{a\theta_i}, A W_{aa} A', W) \quad (18)$$

(ii) Give two scalars b and c , a change $W(a, \theta) \rightarrow bW(a, \theta) + c$ simply multiplies K by b .

(iii) When the cost function K is evaluated with: $W = -\frac{1}{2}(a - \theta \cdot x)^2$, $\|x_i\| = 1$ for all i , and $\|\eta_a\| = 1$ we have

$$K = \kappa^\theta \sum_i |\theta_i|. \quad (19)$$

Then, we have the penalty of θ must be the one in Step 1 of the Algorithm 1, i.e.

$$K(\theta_i, m_a, W_{a\theta_i}, W_{aa}, W_a, W) = \kappa^\theta \sum_i \|\Delta_{\theta_i} \Delta_{\eta_a} W\|.$$

Proposition 1 justifies in some sense Step 1 of the algorithm. We match the basic quadratic targeting of the earlier section, and the model satisfies scale invariance. That leads to the formulation of $\kappa[\theta]$ in Step 1 of the algorithm. ³

³Note that the K function cannot depend on W_a , as this value is generally be 0 in the default policy.

Step 2 is justified, heuristically, by using the idea that penalties for changing one’s representations and penalties for changing one’s action are treated symmetrically. This is why (16) is simply the rewriting of (14) by changing the roles of actions and representations.

The above might be a formal nicety, or perhaps it might reflect something slightly deeper in people’s decision-making: The “basic” algorithm would be given by the penalty (19), and then the mind would simply use the core algorithm after rescaling for the particular units of a situation. That leads the mind to the algorithm (14).

Before enriching the model, we apply it to a concrete problem, so we can better see how it works.

3.2 Application: Quadratic Target Problem

3.2.1 Applying the Sparse BR Algorithm

We detail the application of the model to the quadratic target problem. The value function is

$$W(a, \theta, x, \varepsilon) = \frac{-1}{2} (a - \theta \cdot x - \varepsilon)^2$$

The agent has access to a vector of information x , while ε is not known in advance. θ is the vector of weights to put on x , whose true value is q . The problem is to maximize $\max_a \mathbb{E}^x [W(a, q, x, \varepsilon)]$. Instead, the agent will use $W(a, \theta, x, \varepsilon)$, with θ possibly sparse: $\theta_i = 0$ corresponds to not thinking about dimension i . The DM’s response is as follows (the proof is in the Appendix).

Proposition 2 (*Quadratic Loss Problem*) *In the quadratic optimization problem, the representation is*

$$\theta_i = \theta_i^d + \tau \left(q_i - \theta_i^d, \kappa^\theta \frac{\sigma_a}{\|x_i\|} \right) \quad (20)$$

and the action taken is:

$$a = a^d + \tau \left(\theta \cdot x - a^d, \kappa^a \sqrt{\sum_i \sigma_{\theta_i}^2 v_{x_i}^2} \right) \quad (21)$$

When $\kappa^\theta = 0$, $\theta = q$, and when $\kappa^a = 0$, $a = \theta \cdot x$.

Equation (20) features anchoring on the default value θ_i^d , and a partial adjustment towards the true value q_i : $\theta_i \in [\theta_i^d, q_i]$.

DMs do not deviate from the default iff $|q_i| \|x_i\| < \kappa^\theta \sigma_a$, i.e. when “dimension i cannot explain more than a fraction $(\kappa^\theta)^2$ of the variance of action a ”. *It is the relative importance of*

attribute i in decision a that matters for whether or not the DM will pay attention to attribute i , not its absolute importance in terms, say, of a dollar payoff.

The total amount of attention concerning decision a is the same whether a is very important or less important; this is a correlated from the fact that the model is scale-invariant in W . People will pay attention to say 80% of attributes, whether it is for a small decision (e.g. buying at the supermarket) or a big decision (e.g. buying a car or a house). Some evidence consistent with that is presented by Samuelson and Zeckhauser (1988). I conjecture that it is a good benchmark. It would be good to evaluate this feature empirically.⁴

Equation (21) indicates that, *when there is more uncertainty about the environment, the action is more conservative and closer to the default*: when $\sqrt{\sum_i \sigma_{\theta_i}^2 v_{x_i}^2}$ is higher, a is closer to a^d . In the model, for a given amount of information ($\theta \cdot x$) the power of default is higher when there is more residual uncertainty in the environment. That might be testable in the rich literature on defaults (Madrian and Shea 2001).

We can venture a word about the calibration. As a rough baseline, we can imagine that people will search information that accounts for at least $\xi^2 = 25\%$ of variance of the decision, i.e. if $|q_i| \|x_i\| < \xi \sigma_a$. Then, using (20), we find $\kappa^\theta \simeq \xi$. That leads to the baseline to $\kappa^\theta \simeq 0.5$. By the same heuristic reasoning, we can have as a baseline $\kappa^a \simeq 0.5$.

To conclude, the model generates inattention, inertia, that respond to the local (i.e., for the decision at hand) costs and benefits. Let us now explore the model's consequence in a few applications.

4 Some Applications of the Model

4.1 Choosing n Consumption Goods

We next study a basic, static consumption problem with n goods.

Problem 2 *Suppose that the vector of prices is $p \in \mathbb{R}_{++}^n$, and the utility function is quasi-linear in money. The frictionless decision problem is $\max_{c \in \mathbb{R}^n} u(c) - \lambda p \cdot c$.*

The vector of prices is $p^d + q$, where p^d is the usual price, and q is some change in the price. For instance, in the experimental setup of Chetty, Looney, and Kroft (2009), q could be a tax added to the price. The DM may pay only partial attention to the change in price, and consider that the price of good i is $p(\theta)_i = p_i^d + \theta_i$.

The utility function is quasi-linear in money, but that will be relaxed in section 7.1.1.

⁴One could change that prediction by some choice of κ , though that would take us here too far afield.

We calculate

$$W_{c_i} = u_i - \lambda (p_i^d + \theta_i), \quad W_{c_i c_j} = u_{ij}, \quad W_{c_i \theta_j} = -\lambda 1_{i=j}$$

Hence, the components of the loss matrix are $\Lambda_{ii} = \frac{\lambda^2}{-u_{ii}}$ in two cases: first, if the utility function is separable in the goods ($u(c) = \sum_i u^i(c_i)$), or, for a non-separable utility function, if we apply the algorithm with the “key action” (Section 7.1.3) (the key action for p_i is c_i).

Calling $\sigma_{c_i} = \|\eta_{c_i}\|$, the allocation of attention is:

$$\min_{\theta} \sum_i \frac{\lambda^2 (\theta_i - q_i)^2}{2 |u_{ii}|} + \kappa \lambda |\theta_i| \sigma_{c_i}$$

hence we have

$$\theta_i = \tau \left(q_i, \frac{\kappa |u_{ii}| \sigma_{c_i}}{\lambda} \right) = \tau \left(q_i, \frac{\kappa u_i \left| \frac{c_i u_{ii}}{u_i} \right| \frac{\sigma_{c_i}}{c_i}}{\lambda} \right)$$

and using $u_i = \lambda p_i$, and calling $\psi_i = u_i / (-c_i u_{ii})$ the price-elasticity of demand of good i , we obtain:

$$\theta_i = \tau \left(q_i, \kappa \frac{p_i \sigma_{\ln c_i}}{\psi_i} \right) \quad (22)$$

To go further, we examine the case where preferences are separable, so that the f.o.c. $u_i(c_i) = \lambda p_i$ implies that a change in price dp_i implies $u_{ii} dc_i = \lambda dp_i$, so that $|u_{ii}| \sigma_{c_i} = \lambda \sigma_{p_i}$, hence:

$$\theta_i = \tau (q_i, \kappa \sigma_{p_i}) \quad (23)$$

We see that the price inertia is independent of much of the utility function (e.g., it’s independent of the elasticity of the good): it’s solely dependent on the price of the good.

Equation (22) says that that controlling for the volatility of consumption, inattention is larger for less elastic goods. The intuition is that for such goods, the price is actually a small component of the overall purchasing decision (whose range is measured by $\sigma_{\ln c_i}$). Equation (23) says that to be remarked, a given price change has to be big as a fraction the normal price volatility. It would be good to test those predictions. Chetty, Looney and Kroft (2009) presents evidence for inattention, but do not specifically test a relation like (22) and (23).

4.2 Optimal Monopoly Pricing and a BR-Induced Price Stickiness and Sales

I next study the best response of a monopolist to the behavior of a monopolist facing the BR consumer who has the utility function $u(y, Q) = y + Q^{1-1/\psi} / (1 - 1/\psi)$ for money y and

consumption Q of the good⁵ So if the price is p , the demand is $D(p) = p^{-\psi}$, where ψ is the demand elasticity. The consumer uses the sparse BR algorithm; by the previous analysis, his demand will be:

$$D^{BR}(p) = D(p^d + \tau(p - p^d, \kappa)) \quad (24)$$

where, by (22), $\kappa = \kappa^\theta p_i \sigma_{\ln Q} / \psi_i$. Hence, the consume is insensitive to price changes when $p \in [p^d - \kappa, p^d + \kappa]$.

The monopolist's picks p to maximize profits, $\max_p (p - c) D^{BR}(p)$, where c is the marginal cost (in this section, to conform to the notations of this the optimal pricing literature, c is a marginal cost, while consumption is indicated by Q). The following Proposition describes the optimal pricing policy:

Proposition 3 *With a BR consumer, the monopolist's optimal price is*

$$p(c) = \begin{cases} \frac{\psi c + \kappa}{\psi - 1} & \text{if } c < c_1 \\ p^d + \kappa & \text{if } c_1 \leq c \leq c_2 \\ \frac{\psi c - \kappa}{\psi - 1} & \text{if } c > c_2 \end{cases} \quad (25)$$

where $c_1 = c^d - \sqrt{\frac{2c^d \kappa}{\psi - 1}} + O(\kappa)$ solves equation (47), $c_2 = c^d + \kappa$, with $c^d = (1 - 1/\psi)p^d$. The pricing function is discontinuous at c_1 and continuous elsewhere.

Let us interpret Proposition 3. When $p \in (p^d - \kappa, p^d + \kappa)$, the demand $D^{BR}(p)$ is insensitive to price. Hence, the monopolist won't charge a price $p \in (p^d - \kappa, p^d + \kappa)$: she will rather charge a price $p = p^d + \kappa$. Hence, we get a whole intervals of prices that are not used in equilibrium, and much bunching at $p = p^d + \kappa$. There, price is independent of marginal cost. This is a real "stickiness", and can be a nominal one too. This effect is illustrated⁶ in Figure 2. We see that a whole zone of prices are not used in equilibrium: there is a gap distribution of deviations of prices from the norm. For low enough marginal cost c , prices fall discretely, like a "sale".

The cutoff c_1 is asymmetric. It deviate from the baseline c^d as a square root of the cost.⁷

This simple model seems to account for a few key stylized facts. Prices are "sticky", with a wide range insensitive to marginal cost. This paper predicts "sales": a temporary large fall

⁵Previous work on rational firms and inattentive consumers include L'Huillier (2010), using differently-informed consumers, and Matejka (2010), using a Sims-type entropy penalty. Their model is quite different from the one presented here in both assumptions and results.

⁶The assumed values are $\psi = 6$, $c^d = 7.1$, $\kappa = 0.025p^d$. They imply $p^d = 8.5$, $\kappa = 0.22$, $p(c_1^-) = 7.46$, $p(c_2) = 8.7$.

⁷There is also a more minor effect. For very low marginal cost, consumers don't see that the price is actually too low: they replace p by $p + \kappa$. Hence, reacts less to prices than usual (demand is less elastic), which leads the monopolist to raise is prices. For high marginal cost, consumer replace the price by $p - \kappa$, so their demand is more elastic, and the price is less than the monopoly price.

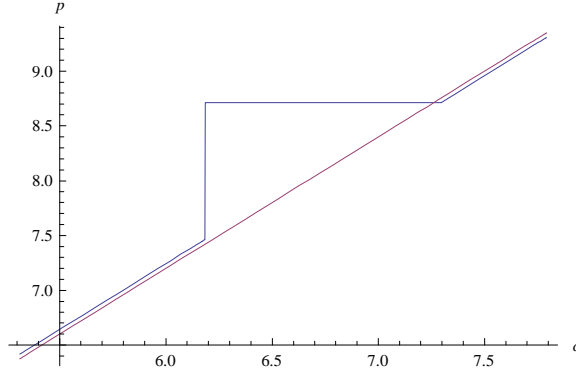


Figure 2: Optimal price as a function of the marginal cost. In the range $(c_1, c_2] = (6.2, 7.3]$ the price is independent of the marginal cost.

in the price, after which the price comes back exactly where it was (if c goes back to $[c_1, c_2]$). This type of behavior documented empirically by Kehoe and Midrigan (2010). In addition, the model says that the typical size of a sales will be $p(c_2) - p(c_1^-)$, i.e., to the leading order

$$p(c_2) - p(c_1^-) = \frac{\sqrt{2\kappa\psi}p^d}{\psi - 1} = \frac{\sqrt{2\kappa^\theta\sigma_{\ln Q}}p^d}{\psi - 1} \quad (26)$$

where we use $\kappa = \kappa^\theta p_i \sigma_{\ln c_i} / \psi_i$ for the last equality.

Hence, the model makes the testable prediction that the gap in the distribution of price changes, and the size of sales, is higher for goods with high consumption volatility, and and that are less price-elastic. The psychological intuition for this is that for those goods, price is a less important factor in the overall purchasing decision. Again, I do not know of no evidence on this.

To close the model, one needs a theory of the default price. In a stationary environment, the simplest is to sa that p^d is the average empirical price,

$$p^d = \mathbb{E} [p(\tilde{c}, p^d)] \quad (27)$$

taking the distribution over the marginal costs \tilde{c} . By the implicit function theorem, for small enough κ and a smooth non-degenerate distribution of costs, there is a fixed point p^d . In the small κ limit and a symmetrical cost distribution f , one can show that $p^d = \frac{\psi}{\psi-1} \bar{c} + \frac{\bar{c}f(\bar{c})\kappa}{2(\psi-1)^2} + o(\kappa)$ with $\bar{c} := \mathbb{E}[c]$ (the calculations are in the online appendix). Hence, the default price is slightly higher than price without BR.

This example illustrates that it is useful to have a tractable model: the sparse-BR algorithm to think about the consequences of bounded rationality in market settings.⁸ Also, those the

⁸For instance much the same analysis will carry over to a closely related setup where consumers are inattentive

sparse BR model was designed to generate inattention, it was not designed to generate price stickiness and sales. Rather, it generates a new theory of those as an unexpected by-product.

4.3 Intertemporal Consumption Problems

4.3.1 Two-Period Consumption Problems

The agent has initial wealth w , and future income y , can consume c_1 at time 1, and invest the savings at a rate R . Hence, the problem is as follows.

Problem 3 (*2-Period Consumption Problem*). *Given an initial wealth w , solve*

$$\max_{c_1} W = u(c_1) + \mathbb{E}[v(y + R(w - c_1))].$$

The income is $y = y_* + \sum_{i=1}^I y_i$: there are J sources of income, and we normalize $\mathbb{E}[y_i] = 0$. Let us study the solution of this problem with the algorithm.

Imperfections of observation The DM observes the income sources sparsely: he uses the model $y^\theta = y_* + \sum_{i=1}^K \theta_i y_i$, the θ_i to be determined. We call $s_1 = w - c_1$ the amount saved at date 1. Let us follow the Sparse BR algorithm. As the default point, we take $\theta^d = 0$. We calculate:

$$W_{c_1} = u'(c_1) - \mathbb{E}[v'(c_2)R], \quad W_{c_1 c_1} = u''(c_1) + \mathbb{E}[v''(c_2)R^2], \quad W_{c_1 \theta_i} = \mathbb{E}[v''(c_2)R y_i]$$

and for the differential operators, $\Delta_{\theta_1} \Delta_{\bar{c}_1} W = \theta_i W_{\theta_1 c_1} \eta_{c_1}$.

Let us proceed with the part related to future income. The loss function is

$$L = \frac{v''(c_2)^2 R^2}{-W_{c_1 c_1}} \sum_i E[y_i^2] (\theta_i - 1)^2$$

and the program (13) is:

$$\min_{\theta_i} L + \kappa^\theta |v''(c_2)R| \sigma_{c_2} \sum_i |\theta_i| \sigma_{y_i}^2$$

Hence, the solution is:

$$\theta_i = \tau(1, \kappa_i) \quad \kappa_i = \kappa^\theta \frac{W_{c_1 c_1} \sigma_{c_2}}{v''(c_2) R \sigma_{y_i}} \quad (28)$$

to the decimal digits of the price, i.e. $D^{BR}(n+x) = D^{BR}(n)$ for n a positive integer and $x \in [0, 1)$. There will be bunching at a price like \$2.99.

To be more analytical, let us take some functional forms, and assume Gaussian noise, $u(c) = -e^{-\gamma c}$ and $v(c) = -e^{-\rho}e^{-\gamma c}$, where γ is the coefficient of absolute risk aversion and ρ the rate of time preference. Because under the default, which has no uncertainty, $v'(c_2)R/u'(c_1) = 1$, and the exponential specification gives $v''(c_2)R/u''(c_1) = 1$, (28) gives:

$$\kappa_i = \frac{\kappa^\theta \sigma_{c_2}}{\sigma_{y_i}},$$

Hence, the agent maximizes, not under $y_2 = y_{2*} + \theta_1 \mu_y + \theta_2 \varepsilon_y$, but under the MIM:

$$y_2 = y_{2*} + \sum_{i=1}^n \tau \left(1, \frac{\kappa^\theta \sigma_{c_2}}{\sigma_{y_i}} \right) y_i \quad (29)$$

Proposition 4 (*2-Period Consumption Model*) *With full maximization of consumption, the time-1 consumption is:*

$$c_1 = \frac{1}{1+R} \left(R w + \delta/\gamma - \gamma \sigma_\varepsilon^2/2 + y_* + \sum_{i=1}^I \theta_i y_i \right) \quad (30)$$

$$\theta_i = \tau \left(1, \frac{\kappa^\theta \sigma_{c_2}}{\sigma_{y_i}} \right)$$

The marginal propensity to consume (MPC) at time 1 out of time-2 wealth, $\partial c_1/\partial y_i$, is:

$$\left(\frac{\partial c_1}{\partial y_i} \right)^{BR} = \left(\frac{\partial c_1}{\partial y} \right)^{ZC} \cdot \tau \left(1, \frac{\kappa^\theta \sigma_{c_2}}{\sigma_{y_i}} \right) \quad (31)$$

where $\left(\frac{\partial c_1}{\partial y} \right)^{BR}$ is the MPC under the BR model, and $\left(\frac{\partial c_1}{\partial y} \right)^{ZC}$ is the MPC under the zero cognition cost model (i.e., the traditional model). Hence, the marginal-propensity to consume is source-dependent.

Proof. Under the expectations induced by θ , we have

$$c_2 = R(w - c_1) + y^e, \quad y^e := y_* + \sum_{i=1}^I \theta_i y_i$$

By maximization, with those beliefs, $E^\theta [e^{-\gamma(c_2 - c_1) - \delta}] = 1$, i.e.

$$-\gamma \left[y^e - \gamma \sigma_\varepsilon^2 + R(w - c_1) - c_1 \right] - \delta = 0$$

i.e. (30). ■

Different incomes sources have different marginal propensity to consume – this is reminiscent

of Thaler’s mental accounts. Equation (31) makes another prediction, consumers pay more attention to sources of income that usually have large consequences, i.e. have a high σ_{y_i} . Extending slightly the model, it is plausible that a shock to the stock market does not affect my disposable income much – hence, there will be little sensitivity to it.⁹

There is a similarity of this model with models of inattention based on a fixed cost of observing information (Duffie and Sun 1990), in particular with the optimal rules of the allocation of attention developed by Abel, Eberly and Panageas (2010), Gabaix and Laibson (2002), and Reis (2006). Because of the fixed cost, in those models the rules are of the type “look up the information every D periods”. From a formal point of view, the present model is more general. Also, the inattention formula (30) holds consumer by consumer, rather than for an consumer that aggregates lots of different consumers – hence the present model is simpler. From a substantive point of view, the range of inactions are different: in the adjustment-every- D -periods model, adjustment will happen. However, in the present model, if the adjustment is always small, then adjustment may never happen. The presence of different models of boundedly rational behavior may help empirical research in that area.

Likewise, we get the following effects (to be typed in a next version of this paper). The MPC of news about y_1 is as follows: high for small changes, low for bigger changes — if the default is “save $s_1^d = \text{fixed}$ ”. If the default is “consumed c_1^* ” (i.e., “save $s_1^d = y_1 - c_1^d$ ”), then MPC is low for small changes, higher for big changes. The model also allows to measure the impact of changes in R : very small. Hence, the measured intertemporal elasticity of substitution, say, will be quite small.

The Euler equation will only hold with the “modified” parameters under \mathbb{P}_θ . Hence, we have

$$\mathbb{E}_\theta \left[\frac{v'(c_2)}{u'(c_1)} R \right] = 1$$

but using the expectation under θ . Note that it features underreaction to future news, especially small future news.

4.3.2 Multiperiod Consumption Problems

One example: say a_t is the amount of money saved at time t . Then

$$W = \sum_{t=1}^{T-1} \beta^t u(y_t - a_t) + v(B), \quad B = \sum_{t=1}^{T-1} R^{T-t} a_t. \quad (32)$$

If the action is c_t , then it is the same as $c_t = y_t - a_t$. Just the default is expressed differently:

⁹In other cases, the default policy might be to consume what’s in one’s wallet, up to keeping a minimum amount, say. Then, the MPC of a dollar bill found on the sidewalk will be 1.

$a_t = s_* y_t$. We calculate:

$$\begin{aligned} W_{a_t} &= -\beta^t u(c_t) + R^{T-t} v'(B) \\ W_{a_t a_s} &= \beta^t u''(c_t) \delta_{st} + R^{2T-t-s} v''(B) \\ W_{a_t y_s} &= -\beta^t u''(c_t) \delta_{st} \end{aligned}$$

So $W_{y_a} W_{a_a}^{-1} W_{a_y}$ is rather messy, because $W_{a_t a_s}$ is non-diagonal. However, calculate the “key action”: for the y_s shock (section 7.1.3), it is just the saving at time s . $t^*(y_s) = s$. Then, with $y_t = \bar{y}_t + \theta \varepsilon_t$

$$(\theta - q)' \Lambda^{diag} (\theta - q) = \sum_t (\theta_t - q_t)^2 \frac{(\beta^t u''(c_t))^2}{\beta^t u''(c_t) + R^{2T-2t} v''(B)} \sigma_{\varepsilon_t}^2$$

Hence, the representation problem is

$$\min_{\theta_t} \frac{1}{2} (\theta_t - q_t)^2 \frac{(\beta^t u''(c_t))^2}{\beta^t u''(c_t) + R^{2T-2t} v''(B)} \sigma_{\varepsilon_t}^2 + \kappa |\theta_t - \theta_t^d| \beta^t |u''(c_t)| \sigma_{a_t}$$

i.e.

$$\min_{\theta_t} \frac{1}{2} (\theta_t - q_t)^2 \frac{1}{1 + R^{2T} (\beta R^2)^{-2t} v''(B) / u''(c_t)} \sigma_{\varepsilon_t}^2 + \kappa |\theta_t - \theta_t^d| \sigma_{a_t}$$

So, if $\beta R^2 < 1$, then the weight in losses falls with time. That means that the DM wants to be precise about early representations, less precise about late representations. There is also the simpler effect that, as one is closer to a given y_t , there is more to adjust because one knows more about y_t .

So the model generates low sensitivity to information about distant future.

4.4 Endowment Effect

We will see that the model generates an endowment effect. Call $a \in [0, 1]$ the quantity of mugs owned, $x \geq 0$ the (random) utility (expressed in a money metric) for having a costless mug, and p the mug price. So the utility is $W(a, x) = a(x - p)$, and the decision problem is $\max_{a \in [0, 1]} W(a, x) = a(x - p)$. Using part 2 of the Sparse BR algorithm (equation 16), we have $\Delta_{\eta_x} W = a \eta_x$, and $\Delta_a \Delta_{\eta_x} W = (a - a^d) \eta_x$, so the problem is:

$$\max_{a \in [0, 1]} a(\mathbb{E}[x] - p) - \kappa^a \|\eta_x\| |a - a^d|$$

where η_x is the uncertainty about x , say $\|\eta_x\| = \sigma_x$.

The solution is simple and yields the willingness to pay (WTP) as well as the willingness

to accept (WTA) for the mug. If $a^d = 0$ (i.e. the agent does not already own the mug), the solution is: buy iff $p \leq WTP = \mathbb{E}[x] - \kappa^a \sigma_x$. If $a^d = 1$ (i.e. the agent already owns the mug), the solution is: sell iff $\mathbb{E}[x] \leq WTA = \mathbb{E}[x] + \kappa^a \sigma_x$. The discrepancy between the two,

$$WTA - WTP = 2\kappa^a \sigma_x \quad (33)$$

is the endowment effect. In contrast, with loss aversion the discrepancy is

$$WTA - WTP = (\lambda - 1) \mathbb{E}[x] \quad (34)$$

where $\lambda \simeq 2$ is the coefficient of loss aversion. (With loss aversion λ , as selling the good creates a loss of $\lambda \mathbb{E}[x]$, while getting it creates a gain of only $\mathbb{E}[x]$).

Hence, this paper's approach predicts that the endowment effect is increasing in uncertain subjective utility (σ_x) of goods.

There is some consistent evidence: for instance, there's no endowment effect for dollar bills, say, which have a known hedonic value. Likewise, professional traders (List 2003) do not exhibit an endowment effect – in this theory, this is because there's a known value to the good.

The next section presents some applications of the model.

4.5 Rebalancing the Portfolio in Flow rather than Stock

We shall see that, in the model, when people think that the stock market is overvalued, they change their stock allocation in the “flow” (invest more in cash and less in new stocks) more than they do in the “stock” (e.g. sell stock in their retirement account and invest the proceeds in bonds).

To see this, say that the retirement account has S in stocks and B in bonds. The agent has an amount y in cash, to be invested either in stocks (s) or bonds (b). The utility is (calling R the random return of stocks):

$$u((S + s)R + Y - S + y - s)$$

We start with an allocation S^d , and s^d . The agent just learned that R is less favorable than he initially thought. What does he do?

Economically, only $S + s$ matters. Which one will change, though? We have

$$\kappa^s = \kappa_0 \|u'' \cdot R\| \|\eta_s\| |s - s_d|, \quad \kappa^S = \kappa_0 \|u'' \cdot R\| \|\eta_S\| |S - S_d|$$

so the f.o.c. is:

$$\mathbb{E}[u'(W) \cdot R] - \kappa_0 \|u'' \cdot R\| (\|\eta_s\| \text{sign}(s - s_d) + \|\eta_S\| \text{sign}(S - S_d)) = 0$$

Hence, if $\|\eta_s\| < \|\eta_S\|$ (which is very likely as the s account is much smaller), the agent first changes the s account, i.e. the flow allocation, rather than the retirement allocation. Given that this preference is strict, it is robust to adding various small costs and frictions.

Going one step further, agents with high cognition costs will rebalance in “flows” and agents with lower cognition costs will rebalance both in stock and in flows. Hence, the model predicts that agents who rebalance in stock are more sophisticated than agents that rebalance by adjusting their flow allocation. If this prediction is verified empirically, it might offer a way to study the “sentiment” of more naive investors: the typical active allocation made in flows, minus the typical active allocation made in stock.

5 Other Features of Representations of the World

5.1 Simplification of Random Variables

5.1.1 Formalism

Consider a random variable Y with values in \mathbb{R}^n . In its MIM, the DM might replace it with a random variable X that might be “simpler” in some sense.

(i) It might have a different, arguably simpler distribution: for instance, we could replace a continuous distribution with a 1-point distribution (e.g. $X = \mathbb{E}[Y]$ with probability 1), or with a two-point distribution $X = \mathbb{E}[Y] \pm \beta$, for some β . We could even have X be a certainty equivalent of Y .

(ii) It might have independent components. For example, we could have $X_i \stackrel{d}{=} Y_i$, but the components $(X_i)_{i=1\dots n}$ are independent, while the components $(Y_i)_{i=1\dots n}$ are not.

To formalize (i), call F and G the CDF of X and Y respectively. Then, $U = G(Y)$ has a uniform $[0, 1]$ distribution, and we can define $X = F^{-1}(U)$, with the same U so that X and Y are maximally affiliated.

To formalize (ii), it is useful to use the machinery of copulas. For an n -dimensional, let us write $Y = (G_1^{-1}(U_1), \dots, G_n^{-1}(U_n))$ with U_1 have the copula $C(u_1, \dots, u_n)$, so that $\mathbb{E}[\phi(Y)] = \int \phi(G_1^{-1}(u_1), \dots, G_n^{-1}(u_n)) dC(u_1, \dots, u_n)$. In the simplified distribution, the marginals G_i^{-1} could be changed, and the copula could be changed. To express X , we could have $X = (G_1^{-1}(U'_1), \dots, G_n^{-1}(U'_n))$, where the U'_i might have the copula of independent variables, $C^\theta(u_1, \dots, u_n) = u_1 \cdots u_n$, of some intermediary copula. If we wish to have X_i 's marginals

simpler than Y_i 's, like in (i), we can set $X = (F_1^{-1}(U'_1), \dots, F_n^{-1}(U'_n))$ for some F_i .

Eyster and von Weizsacker (2010) present experimental evidence for correlation neglect, i.e. the use of simplification (ii). The next example illustrates the possible relevance of simplification (i).

5.1.2 Application: Acquiring-a-company Game

Samuelson and Bazerman (1985) devised a ingenious problem.

Problem 4 (*Acquiring-a-company*) *The company is worth $X \sim U[0, 100]$ to Ann, and worth $1.5X$ to you (you're a better manager than Ann). You can make a take-it-or-leave-it offer a to Ann, who knows X . Which offer do you make?*

In addition, the experimental set up makes sure that “Ann” is a computer, so that its answer can be assumed to be rational. Experimentally, subjects respond with a mode around 60, and a mean around 40 (Charness and Levin 2009). However, the rational solution is $a = 0$. This is a case of extreme asymmetric information.

The objective payoff is:

$$\pi(a) = \mathbb{E} \left[\left(\frac{3}{2}X - a \right) 1_{X \leq a} \right] \quad (35)$$

Let's see how to state the MIM. For simplicity, we normalize the maximum X to 1, $X \sim U[0, 1]$. We will see how, if the agent uses a simpler representation of probabilities, we account for the non-zero experimental value. This is a different explanation from available explanations (Eyster and Rabin 2005, Crawford and Iriberri 2007), which emphasize assuming that the other player is irrational, but the DM is rational. However, there is no “other player” in this game, as it is just a computer, and then those models predict a bid of 0 (Charness and Levin 2009).

In the simplest representation, $X^\theta = \frac{1}{2}$: the agent forms a model of the situation by simplifying the distribution, replacing it with a distribution with point mass $X = 1/2$. Then, the best response is $a = 1/2$. This is not too far from the empirical evidence.

In a richer MIM, let us replace the distribution with a 2-point distribution, $X^\theta = 1/2 - \theta$ with probability $1/2$, $X = 1/2 + \theta$ with probability $1/2$, for some $\theta \in [0, 1/2]$ (we leave it to be an empirical matter to see what θ is – the same way it is an empirical matter to see what the local risk aversion is). Given this model, the agent solves for (35).

Proposition 5 *In the acquiring-a-company problem, the Sparse BR bid by the decision-maker is:*

$$a^* = \begin{cases} \frac{1}{2} + \theta & \text{if } \theta \in [0, \frac{1}{6}] \\ \frac{1}{2} - \theta & \text{if } \theta \in (\frac{1}{6}, \frac{1}{2}] \end{cases} . \quad (36)$$

Proof. It is clear that the optimal solution a belongs to $\{0, 1/2 - \theta, 1/2 + \theta\}$. If the offer is $a = 1/2 - \theta$, the offer is accepted only if $X = 1/2 - \theta$ (in the MIM), so:

$$\pi^{M2} \left(\frac{1}{2} - \theta \right) = \frac{1}{2} \cdot \left(\frac{3}{2} \left(\frac{1}{2} - \theta \right) - \left(\frac{1}{2} - \theta \right) \right) = \frac{1}{8} - \frac{\theta}{4}$$

If the offer is $a = 1/2 + \theta$ the buyer gets the plant for sure, which has a value to him of $\frac{3}{4}$ in expectation, so:

$$\pi^{M2} \left(\frac{1}{2} + \theta \right) = \frac{3}{4} - \left(\frac{1}{2} + \theta \right) = \frac{1}{4} - \theta$$

The two profits $\pi^{M2} \left(\frac{1}{2} - \theta \right)$ and $\pi^{M2} \left(\frac{1}{2} + \theta \right)$ are the same if and only if $\theta = 1/6$. So, the optimal decision is as announced in the Proposition. Hence, the maximum paid is $\frac{1}{2} + \frac{1}{6} = \frac{2}{3}$.

■

Likewise, if the agent uses a 3-point distribution at $1/2 + k\theta$, $k \in \{-1, 0, 1\}$, then the optimal offer a^* is: $1/2 + \theta$ for $\theta \in [0, \frac{1}{6}]$, $1/2$ for $\theta \in [\frac{1}{6}, \frac{1}{4}]$, and $1/2 - \theta$ for $\theta \in (\frac{1}{4}, \frac{1}{2}]$. Hence, the predictions are quite similar.¹⁰

On the other hand, the model does not explain part of the results in the Charness and Levin (2009) experiments. In a design where the true distribution of X is 0 or 1 with equal probability, the rational choice is $a = 0$. However, subjects' choices exhibit two modes: one very near 0, another around $a = 1$. The model explains the first mode, but not the second one. It could be enriched to account for that additional randomness, but that would take us too far afield. One useful model is the contingency-matching variant of Section 10.1.3: with equal probability, the DM predicts that the outcome will be 0 and 1, and best-responds to each event, so plays 0 or 1 with equal probability. Hence, reality seems to be reasonably well accounted for by a mixture of the basic model and its contingencies-matching actions.

All in all, the model does a rather nice job at describing behavior in the basic acquiring-a-company game, even though it doesn't account for all the patterns in the other variants.

5.2 Dictionaries and Stereotypical Thinking

One particular interpretation of the θ is potentially interesting. We could have a “dictionary”: $(\theta_i)_{i \in I}$ for some index set I , and $\theta_i \in \Theta_i$ for a some set Θ_i . The resulting representation is:

$$X(\theta) = \sum_{i \in I} x_i(\theta_i)$$

¹⁰I note that the empirically correct prediction $a^* > 1/2$ holds only if $\beta \leq 1/6$. In particular, that means that $\beta < stdev(X) = \frac{1}{2\sqrt{3}} = 0.29$, which could have been a normatively appealing benchmark. Also, $\beta < AbsDev(X) = 1/4 (= \mathbb{E}[|X - \mathbb{E}[X]|])$.

Note that the dictionary might be “redundant,” i.e. the $x_i(\theta_i)$ need not form a basis.

For instance, take a geometrical example and the plane \mathbb{R}^2 . We could have: $x_1(\alpha, \beta, R)$ the circle with center (α, β) and radius R (the index is in \mathbb{R}^3); $x_2(\alpha, \beta, \alpha', \beta')$ a square starting with two “diagonal” edges (α, β) and (α', β') (the index is then in \mathbb{R}^4). The total figure is the sum of all those primitive figures. We describe a picture from the basic constituents.

In a more social setting, we could have x a n -dimensional vector of attribute such as profession, nationality, income, social background, ethnicity, gender, height etc. Then, the primitive words in the dictionary could be x_{Eng} for a stereotypical engineer, x_{Asian} for an Asian person, etc.

The key is that it’s easy (sparse) to think in terms of “ready-made” categories, but harder (less sparse) to think in terms of a mix of categories. For instance, suppose that the space of attributes is $X = (y_1, y_2)$, where y_1 is how good the person is at mathematics, and y_2 how good she is at dancing. Say that there’s “type” engineer, with characteristics $x_{\text{Eng}} = (+8, -3)$: i.e. engineers are quite good at math, but are rather bad dancers (on average). Take a person called Johanna. First, we’re told she’s an engineer, and the first representation is $x_J = x_{\text{Eng}}$. Next, we’re told she’s actually a good dancer, and with level +4 in dancing. Her characteristics is $x_J = (8, 4)$. How will she be remembered? We could say $x(\theta) = x_{\text{Eng}} + 7x_{\text{Dancer}}$, but such a representation is rather costly. Hence, the information “good dancer” may be discarded, and only x_{Eng} will be remembered. The “stereotype” of the engineer eliminates the information she’s a good dancer.

More precisely, suppose that one wishes to maximize $U = -(a_1 - x_1)^2 - \beta(a_2 - x_2)^2$, i.e. have a good model of the person, with a weight β on the dancing abilities. We start from $x^d = X_{\text{Eng}} = (8, -3)$, and plan to change (x_1^d, x_2) (it is clear that the first dimension need not change). Applying the algorithm, we have $\max_{\theta} -\frac{1}{2}\beta(a_2 - \theta_2)^2 - \kappa\sigma_{a_2}\beta|\theta_2 - x_2^d|$, hence using Lemma 1, $x_2 = x_2^d + \tau(x_2^q - x_2^d, \kappa\sigma_{a_2})$, i.e.

$$x_2 = -3 + \tau(7, \kappa\sigma_{a_2})$$

Hence we get a partial adjustment, with x_2 between the stereotypical level of dancing (-3) and Johanna’s true level (4).

Hence, a model of sparsity-seeking thinking with a dictionary would be the following. Given a situation x , find a sparse representation that approximates x well, e.g. find the solution to:

$$\min_{(\theta_i)_{i \in I}} \left\| \sum_{i \in I} x_i(\theta_i) - x \right\|_2^2 + \sum_i \kappa_i |\theta_i - \theta_i^d|$$

Then, people will remember $x(\theta) = \sum_{i \in I} x_i(\theta_i)$, rather than the true x . That generates a

simplification of the picture, using simple traits. The above may be a useful mathematical model of the categorization. For instance, we might get a model of “first impressions matter”. The first impression determines the initial category. Then, by the normal inertia in this model, opinions are adjusted only partially.

It is also clear that it is useful to have a dictionary of such archetypes: they make thinking, or at the very least, remembering, sparser. One may also speculate that the education and life events give the DM new elements in his dictionary.

6 Multiple Players

6.1 Model Statement

For multiple players, we glue the basic 1-person algorithm to the k -level thinking models of Stahl (1998) and Camerer, Ho, and Chong (2004). They assume perfect rationality for the DM, whereas here the DM is boundedly rational. Also, we express things here in a fairly general dynamic framework.

There are K types of players, indexed $k = 1 \dots K$, with mass m_k . We start with a vector of default policies a_k^d , representations θ_k^d and value function V_k^d . Depending on the context, default policies could be very random, e.g. “pick each action with equal probability.”

Algorithm 2 (*Game with Multiple Players*) *In an K -player game, the algorithm is the following for player 1:*

Step 1. Apply the Sparse BR algorithm to the function:

$$W(a_t, (a_{kt}^d)) = u(a_t, (a_{kt}^d)_{k=1 \dots K}) + \beta V^d(S_{t+1}(a_t, (a_{kt}^d)_{k=1 \dots K}, x_{t+1}))$$

without flexibility in the representation of the world, but with a cost of action equal to $\kappa^a = \kappa$ with probability π , and $\kappa^a = \infty$ with probability $1 - \pi$. That is, a fraction π of DMs do a κ^a -best response, or while a fraction $1 - \pi$ do not deviate from the default policy. Call those policies a_k^1 and a_k^0 , respectively.

Step 2 (round $R = 2$). Update the model of the world: say that there are $2K$ types, (k, r) , $r = 0$ or 1 . Model

$$a_{k,1}^\theta = (1 - \theta_{k,1}) a_k^0 + \theta_{k,1} a_k^1$$

Then, apply the Sparse BR algorithm to the function: W , using the plans that a mass m_k of players will play $a_{k,1}^\theta$ with probability π , and a_k^0 with probability $1 - \pi$. Use the same dichotomy of action maximization, assuming that players who “stopped thinking” in round 1 continue to stop thinking. Call a_k^2 the policy of the players who went through two rounds of thinking.

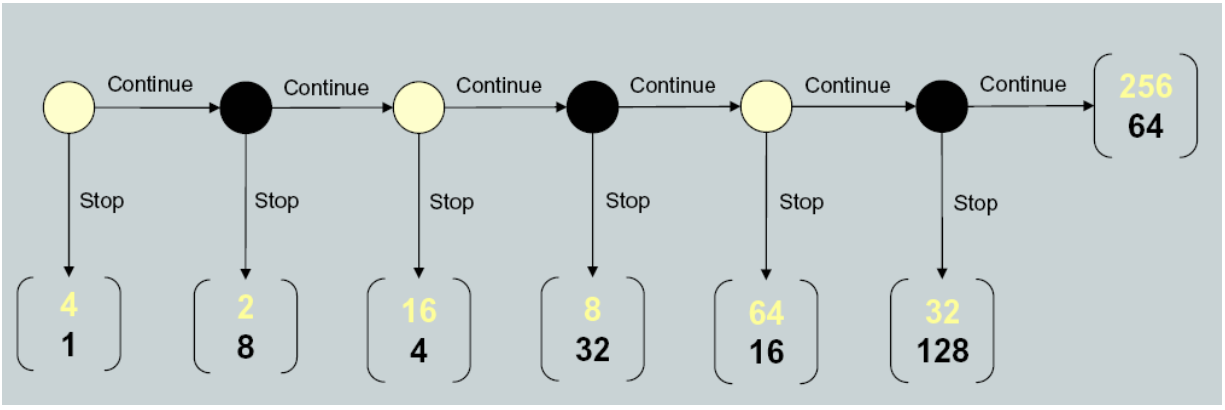


Figure 3: A centipede game. Reproduced from Palacios-Huerta and Volij (2009).

Step 2' (round $R \geq 2$). Update the model of the world, say that there are RK types (k, r) , $r = 0 \dots R - 1$. Model the actions as:

$$a_{k,r}^\theta = (1 - \theta_{k,r}) a_k^0 + \theta_{k,r} a_k^r$$

and proceed as in Step 2. Call a_k^R the resulting action for players who would go through R rounds of reasoning.

Step 3. Iterate on the round R until no player updates his action, or indefinitely $R = \infty$.

In the end, the policies are made of $R_{\max}K$ types $(k, r)_{r=0, \dots, R_{\max}-1}$ with mass $m_k (1 - \pi) \pi^r$, and actions a_k^r .

The algorithm is a little complicated, but in many situations it is easy to use. First, in a 1-player game, it is just the previous Algorithm 1. Second, in one-shot games, when agents who do think have zero cognition costs ($\kappa^a = 0$ and $\kappa^\theta = 0$), the model is identical to the Cognitive Hierarchy Model (Camerer, Ho, and Chong 2001), which has proven to be a useful benchmark. However, it imbeds the Cognitive Hierarchy idea in a model where individual agents maximize imperfectly, and also in dynamic programming situations.

6.2 Applications

6.2.1 Centipede Game

Figure 3 shows a centipede game (Rosenthal 1981), reproduced from Palacios-Huerta and Volij (2009). The game-theoretic prediction is that player 1 stops and gets 4, but in practice very few non-professional players do that.

We assume that the default action is “randomize at each node.”¹¹ Under that scenario,

¹¹We could have a more robust rule, e.g. “randomize at each node, but with a weight 0 on the nodes that

player 1 considers his payoff. A simple calculation shows that the expected payoff of continuing is 19, while the payoff from stopping is 4. Hence, player 1 continues. Indeed, at all nodes, players continue.

The explanation of the centipede game shares similarities to that of McKelvey and Palfrey (1992). In their interpretation, the opponent can be selfish with some probability q , or altruist (and always continue) with some probability $1 - q$. Hence, their model still requires backward induction, using this structure. However, in the present model, only forward induction is needed. Hence, the model is simpler to use in the centipede situation. Also, depending on the costs and benefits of thinking, the models will make different predictions (a future study, beyond the scope of this paper, would be required to do justice to this issue). In addition, Algorithm 2 is fairly generally available.

6.2.2 Dollar Auction Game

The dollar auction game (Shubik 1971) is an amusing and enlightening game. There is an amount of money to auction, say a $D = \$20$ bill. There are two players (for simplicity in this paper) who can participate in an ascending auction with increments of \$1. The person who drops out, and has bid b , pays b . The person who stayed and bid $b' = b + 1$ gets the bill and pays b' , for a total payoff of $D - b'$. The initial bid, if any, is at an arbitrary level.

In practice, one gets an initial bid $b < D$ (say $b = 1$), and then the game escalates without limit (or, until a limit imposed by the instructor is reached). The paradox is that it always seems better to go “one more step” even though the expected payoff is infinitely negative. That is not the equilibrium predicted by orthodox game theory, which is the following: someone bids $b = D$, and the other player doesn’t bid.

The model delivers the classroom result. Suppose that the MIM is that people will continue with probability π and stop with probability $1 - \pi$ (as a baseline, $\pi = 1/2$). In the model, the right thing to do is to best-respond given those future events. To analyze this formally, consider $V(b)$, the expected payoff of continuing with a bid b . It is easy to show that it is equal to:¹²

$$V(b) = D - b - \frac{2\pi}{1 - \pi} \tag{37}$$

can lead to a payoff less than M for some M .”

¹²We have the Bellman equation

$$V(b) = (1 - \pi)(D - b) + \pi \max(V(b + 2), -b - 2)$$

Indeed, with probability $1 - \pi$, the other player drop out with probability $1 - \pi$, then the payoff is $D - b$. With probability π he continues, then the player chooses the best option between dropping out (which yields $-b - 2$) or continuing (which yields $V(b + 2)$). To solve the Bellman equation, we seek a solution of the type $V(b) = Ab + B$, and solve for A and B by plugging that solution in the Bellman equation. This yields (37).

It is better to continue and bid $b + 2$ rather than give up iff $V(b + 2) \geq -b$, i.e. iff $D \geq \frac{2\pi}{1-\pi}$. This is the case for instance if $D = \$20$ and $\pi = 1/2$. Indeed, for a wide range of parameters agents will continue forever. They will stop only if they see that the probability of continuing is $\pi > D/(2 + D)$, which is 0.91 when $D = \$20$.

7 Complements and Discussion

7.1 Some Variants on the Model

This subsection indicates some small variants to Sparse BR Algorithm that may be useful in some situations. The reader may wish to skip this section in the first reading.

7.1.1 Model with Constraints

We now come back to the presentation of the model, with a more general framework (which includes consumption problems), that has a number K of constraints.

$$\max_a \mathbb{E}^\varepsilon [u(a, \theta, x, \varepsilon)] \text{ subject to } B^k(a, \theta, x) \geq 0 \text{ for } k = 1 \dots K \quad (38)$$

For instance, B^1 could be a budget constraint, $B^1 = y - p \cdot c$.

We will use the methods of Lagrange multipliers to formulate an algorithm.

Algorithm 3 (*Constrained-Sparse BR Algorithm*) *To solve the problem (38), the agent uses the following three steps.*

1. *Transformation into an unconstrained problem. At θ^d , solve for the problem. Pick the Lagrange multiplier $\lambda \in \mathbb{R}^k$ such that the solution is:*

$$\max_a \mathbb{E}^\varepsilon [u(a, \theta^d, x, \varepsilon)] + \lambda \cdot B(a, \theta^d).$$

2. *BR-Solve the new, unconstrained problem. Use the Sparse BR Algorithm 1 for the value function*

$$W(a, \theta, x, \varepsilon) := u(a, \theta, x, \varepsilon) + \lambda \cdot B(a, \theta, x).$$

That returns a representation θ and an action a .

3. *Adjustment to take the constraints fully into account. Call a^d a default action that satisfies the constraints, and $a(\mu) = a^d + \mu(a - a^d)$. Pick the real μ that satisfies $\max_\mu \mathbb{E}^\varepsilon [u(a(\mu), \theta, x, \varepsilon)]$ subject to the constraints. The chosen action is $a(\mu)$.*

For instance, in the study of section 4.1, if the utility problem is $\max_{c \in \mathbb{R}^n} u(c)$ s.t. $y - p \cdot c \geq 0$, the new Step 1 is to we pick the Lagrange multiplier λ that corresponds to the problem: $\max_c u(c) + \lambda (y - p^d \cdot c)$. Then, we define:

$$W(c, \theta) = u(c) + \lambda \left(y - \sum_i (p_i^* + \theta_i) c_i \right) \quad (39)$$

This gives us a quasi-linear utility function, with linear utility for residual money. The Step 2 is as in section 4.1. In step 3 (applied with $c^d = 0$), the DM picks a consumption bundle $c = \mu \cdot c(\mathbf{p} + \theta, y)$, where μ ensures the budget constraint, $(p + q) \cdot c = y$.

7.1.2 Discrete sets, Non-differential Operators

Sometimes (e.g., when the space underlying a is not continuous) it is useful to replace the differential operators used in Algorithm 1 by their non-differential counterpart (the superscript F is a short-hand for “finite”):

$$\begin{aligned} (\Delta_{\theta_i}^F f)(\theta) &= f(\theta_i, \theta_{-i}) - f(\theta_i^d, \theta_{-i}), & (\Delta_{a_i}^F f)(\theta) &= f(a_i, a_{-i}) - f(a_i^d, a_{-i}) \\ (\Delta_{\eta_\theta}^F f)(\theta) &= f(\theta + \eta_\theta) - f(\theta), & (\Delta_{\eta_a}^F f)(a) &= f(a + \eta_a) - f(a) \end{aligned}$$

How to define “ $a + \eta_a$ ” when the action space A is finite? Assume that space A comes equipped with a distance $d(a, a')$: for instance, if $A = \{1, \dots, n\}$ ordered in \mathbb{N} , $d(a, a') = |a - a'|$, and if A is just a set of options with no clear metric (e.g., 4 options with no particular spatial ordering), then we can have $d(a, a') = 1_{a \neq a'}$. Then, “ $a + \eta_a$ ” with a probability proportional some decreasing function of $d(a, a')$, for instance $e^{-\beta d(a, a')}$ for some β .

Likewise, sometimes (e.g. when dealing with functions with discrete support) to have a non-differential version of the Λ matrix. A simple device is to consider values $a^*(\theta)$ and set:

$$\Lambda_{ii} = \frac{1}{(\theta_i - q_i)^2} \mathbb{E}^{x, \varepsilon} [u(a^*(\theta_i, q_{-i}), q, x, \varepsilon) - u(a^*(q), q, x, \varepsilon)] \quad (40)$$

where $a^*(\theta)$ is the optimum under the model parametrized by θ .

7.1.3 Simplifying the realism loss parameter Λ

The “key action” simplification for Λ The following simplification is often useful. For Λ , use

$$\Lambda^{diag} = diag(\Lambda_{11}, \dots, \Lambda_{nn}), \quad \Lambda_{ii} = \max_k \frac{-W_{\theta_i a_k}^2}{W_{a_k a_k}} \quad (41)$$

The intuition is the following. For each dimension θ_i , pick the “key action” that is related to it: the one with the maximum $\frac{-W_{\theta_i a_k}^2}{W_{a_k a_k}}$, in virtue of (43). The term Λ^{diag} is simple to calculate,

and doesn't involve the matrix inversion of the general Λ in (12).

Averaging In the baseline model Λ is evaluated at the default action and representation. We could extend that by averaging around the baseline. For instance, define $\Lambda(a, \theta, x) = -W_{a\theta}W_{aa}^{-1}W_{a\theta}$ and

$$\Lambda = \mathbb{E} [\Lambda(a^d + \eta_a, \theta^d + \eta_\theta, x)] \quad (42)$$

where the expectation is over η_a, η_θ , and x . So we add noise around a^d and θ^d .

Application: if we use the default action (no saving), there is no impact of the interest rate, the simple Λ is 0. But with the average, the agent will see that for some other policies (non-zero saving) the interest rate does matter.

7.1.4 Parameter-specific Complexity Cost

Different sources could have different complexity. This is easy to represent as $\kappa[\theta] = \sum_i \kappa_i^\theta \|\Delta_{\theta_i} \Delta_{\eta_a} W\|_\alpha$ where costlier sources have a higher κ_i^θ .

7.1.5 Enrichments in $\kappa[a]$

Enrichment via loss aversion One interesting enrichment is to use a loss-aversion based penalty that penalizes negative outcomes but not positive ones. Denote $x^- = \max(-x, 0)$, i.e. $x^- = -x$ for $x < 0$ and 0 for $x \geq 0$. Call Δ_- the ‘‘loss aversion’’ operator, $(\Delta_- f)(x) = (f(x))^-$. Instead of the original formulation (16), $\kappa[a] = \kappa^a \sum_i \|\Delta_{a_i} \Delta_{\eta_\theta} W\|_\alpha$, we could have, for a complexity parameter $\kappa^{a,-}$:

$$\kappa[a] = \kappa^{a,-} \sum_i \|\Delta_- \Delta_{a_i} W\|$$

This operator Δ_- may be useful, first, because loss aversion seems important in many parts of economic psychology. Also, it is serviceable in the (relatively rare) cases where a gamble is offered with no downside. To see this, take the problem where the agent can pick a quantity $a \in [0, 1]$ of a gamble g with non-negative support, i.e. obtain utility $u(ag)$. It is clear that, whatever the complexity of g , by domination, picking $a = 1$ is the right thing to do. This is missed by the basic algorithm, but is detected with the loss aversion operator: normalizing $u(0) = 0$,

$$\kappa[a] = \kappa^{a,-} \|\Delta_- \Delta_a W\| = \kappa^{a,-} E[(u(ag) - u(0))^-] = 0$$

because $u(ag) - u(0) \geq 0$ almost surely. Then, it is clear that there is no penalty for complexity.

We can also mix and match, and replace (16) by

$$\kappa[a] = \kappa^a \sum_i \|\Delta_- \Delta_{a_i} \Delta_{\eta_\theta} W\|_\alpha + \kappa^{a,-} \sum_i \|\Delta_- \Delta_{a_i} W\|$$

This is adding a “loss aversion” operator to the previous operators. It seems that in many situations it is not worth bothering the loss aversion operator Δ_- , which adds some algebraic complexity, but it is good to have it available when “dominations” patterns are important.

Finally, the DM might restrict himself to a parametrization of the actions. For instance, if the underlying action is $A = (A_1, \dots, A_T)$, and A_t is the savings rate at time t , we can have $A_t(a) = a_0 + a_1 t$, a savings rate that depends in an affine way on age, where (a_0, a_1) is a 2-dimensional parametrization of the agent’s savings rate.

7.2 Links with Themes of the Literature

7.2.1 Links with Themes in Behavioral Economics

In this section, I mention the ways the Sparse BR approach meshes with themes in behavioral economics: it draws from them, and is a framework to think about them.

Anchoring and adjustment The model exactly features anchoring and adjustment for expectations and decisions: the anchor is the default MIM θ^d and action a^d , the adjustment is dictated by the circumstances.

Power of defaults Closely related to anchoring and adjustment, it has now been well established that even in the field default actions are very often followed (Madrian and Shea 2001, Carroll et al. 2009). This model prominently features that stylized fact.

Rules of thumb Rules of thumb are rough guides to behavior, such as “invest 50/50 in stocks and bonds,” and “save 15% of your income.” They are easily modeled as default actions. The advantage is that the model will generate deviations from the rule of them (the default action) when the circumstances call for it with enough force: for instance, if income is very low, the agent will see that current marginal utility is very high, and he should save less.

Temptation vs BR The present model is about bounded rationality, rather than “emotions” such as hyperbolic discounting (Laibson 1997) or temptation. Following various authors (e.g. O’Donoghue and Loewenstein 2005, Fudenberg and Levine 2006, Brocas and Carillo 2008), we can imagine an interesting connection, though, operationalized via defaults. Suppose that “system 1” (Kahneman 2003), the emotional and automatic system, wants to consume now. This could be modeled as saying that System 1 resets the default action to high consumption now (it likely will also shift the default representation). System 2, the cold analytical system, could be modeled as what the current algorithm does. It partially overrides the default when

cognition costs are low, but will tend to follow it otherwise. So, while many papers have focused on modelling “system 1,” this paper attempts at modelling “system 2”.

Mental Accounts Some of the above has a flavor of “mental accounts.” Take the basic wine example by Thaler: the wine was purchased for \$20, now it is \$80. Rationally, people should sell that bottle. However, the default is to consume it. So, if cognition costs are high enough, people will just follow that default. On the other hand, the effect is limited in the model: if the bottle was worth \$8,000, say, people would sell it.

Availability The theory is silent about the cost κ_i of each dimension: in the benchmark model they are the same. It is plausible that more “available” dimensions will have a lower κ_i . For instance, availability is greater when a variable is large, familiar, and frequently used.

First-order risk aversion The model generates first-order aversion, as seen in section 4.4.

$1/n$ heuristics This heuristic (Bernatzi and Thaler 2001, Huberman and Jian 2006) is to allocate an amount $1/n$ when choosing over n plans, independently of the plans’ correlation: for instance, the agent allocates $1/3, 1/3, 1/3$, no matter whether the offering is one bond fund and two stocks funds or the offering is one stock fund and two bond funds. We get that in the model by using the “simplification of variables.” The simplification is that the variables are treated as independent rather than correlated.

7.2.2 Links with Other Approaches to Bounded Rationality

This paper is yet another attempt in a long series of attacks on the polymorphous problem of bounded rationality: see Conslík (1996) for a survey. Some put the accent on learning (Sargent 1993), a theme that could be merged with the current paper. Some model people as finite automata (Rubinstein 1998), an interesting idea that nonetheless is hard to implement in a tractable way. In contrast to some papers, the accent here is on models that can be used directly in economics.

This paper also links to a literature modelling inattention. Some is with fixed cost (Duffie and Sun 1990, Gabaix and Laibson 2002, Reis 2006), some with an entropy-related cost (Sims 2003). This paper, in contrast, recommends the ℓ_1 penalty for sparsity. This seems to be a novel import in modelling bounded rationality, even though it has been quite useful in the applied mathematics literature cited above.

8 Conclusion

This paper proposes a tractable model with some boundedly rational features. No doubt, it could and should be greatly enriched. For instance, it is silent about some difficult operations like Bayesian (or non-Bayesian) updating and learning (see Gennaioli and Shleifer 2010 for recent progress in that direction) and memory management (Mullainathan 2002). Even though it can be applied to situations with several dates, its essence is still quite static. However, despite these current limitations, given its tractability and fairly good generality, it might be a useful point of departure to think about the impact of bounded rationality in economic situations.

9 Proof Appendix

Loss from Lack of Realism What is the DM's loss when he makes an approximation θ while the truth is q ? For a given x , define $W(a, \theta) = \mathbb{E}^\varepsilon [W(a, \theta, x, \varepsilon)]$ and $a(\theta) = \arg \max_a W(a, \theta)$. The loss is $L = W(a(\theta), q) - W(a(q), q)$. Let us approximate in the limit of small losses. As a solves $W_a(a, \theta) = 0$, the implicit function theorem gives $\delta a = -\frac{W_{a\theta}}{W_{aa}} \delta \theta$, so the loss is:

$$L = W_a \delta a + \frac{1}{2} W_{aa} (\delta a)^2 = 0 + \frac{1}{2} W_{aa} \cdot \left(-\frac{W_{a\theta}}{W_{aa}} \delta \theta \right)^2 = \frac{-W_{a\theta}^2 (\delta \theta)^2}{2W_{aa}}$$

By the same reasoning, in n dimensions the loss is

$$L = \frac{1}{2} (\theta - q)' \Lambda (\theta - q), \quad \Lambda = -W_{\theta a} W_{aa}^{-1} W_{a\theta} \quad (43)$$

Proof of Proposition 1. We need the following Lemma.

Lemma 2 Consider positive integers n, p , and a function $f : \mathbb{R}^n \times \mathbb{R}^{n \times p} \rightarrow S$, for some set S , such that for all $x \in \mathbb{R}^n, y \in \mathbb{R}^{n \times p}$, and $A \in \mathbb{R}^{n \times n}$, $f(Ax, y) = f(x, A'y)$. Then, there exists a function $g : \mathbb{R}^p \rightarrow S$ such that $f(x, y) = g(x'y)$.

Proof. Define $e_1 = \begin{pmatrix} 1 \\ 0_{n-1} \end{pmatrix}$ and for a row vector $z \in \mathbb{R}^p$, $g(z) := f(e_1, e_1 z)$. We have:

$$\begin{aligned} f(x, y) &= f(xe_1' e_1, y) \text{ as } e_1' e_1 = 1 \\ &= f(e_1, e_1 x'y) \text{ using the assumption with } A = xe_1' \\ &= g(x'y). \end{aligned}$$

■

Hypothesis (ii) implies that K is independent of the point estimate of W (as opposed to its derivatives): it can be written $K(\theta_i, \eta_a, W_{a\theta_i})$ for some function K (by a minor abuse of notation).

We use the invariance by reparametrization λ_1 in hypothesis (i), and apply Lemma 2 to $K(\theta_1, W_{a\theta_1}, Z_1)$, where Z_1 represents the other arguments. This implies that we can write $K(\theta_1, W_{a\theta_1}, Z_1) = K(\theta_1 W_{a\theta_1}, Z_1)$, for a new function K . Proceeding the same way for $(\theta_i, W_{a\theta_i})$ for $i = 2 \dots n$, we see that we can write $K = K(\eta_a, (\theta_i W_{a\theta_i})_{i=1 \dots n})$. We next apply Lemma 2 to

$x = \eta_a$ and $y = (\theta_i W_{a\theta_i})_{i=1\dots n_\theta}$. It implies that we can write:

$$K = k((\eta_a \cdot W_{a\theta_i} \theta_i)_{i=1\dots n})$$

for some function k .

Finally, to determine k , we use hypothesis (iii), which indicates that whenever $\|x_i\| = 1$, and $\|\eta_a\| = 1$, $k((\eta_a x_i \theta_i)_{i=1\dots n}) = \sum_i |\theta_i|$. By homogeneity, when $\|x_i\|$ and $\|\eta_a\|$ are nonzero, define $\hat{x}_i = x_i / \|x_i\|$, $\hat{\eta}_a = \eta_a / \|\eta_a\|$ and $\hat{\theta}_i = \theta_i \|x_i\| \|\eta_a\|$. Then,

$$\begin{aligned} k((\eta_a x_i \theta_i)_{i=1\dots n}) &= k\left(\left(\hat{\eta}_a \hat{x}_i \hat{\theta}_i\right)_{i=1\dots n}\right) \\ &= \sum_i \left| \hat{\theta}_i \right| = \sum_i \|\eta_a x_i \theta_i\| \end{aligned}$$

so in general,

$$K = \sum_i \|\eta_a x_i \theta_i\| = \sum_i \|\Delta_{\theta_i} \Delta_{\eta_a} W\|.$$

Proof of Proposition 2 *Step 1: Representation.* We calculate

$$\begin{aligned} W_a &= -a + \theta \cdot x + \varepsilon, & W_\theta &= x(a - \theta \cdot x - \varepsilon) \\ W_{a\theta} &= x, & W_{aa} &= -1 \end{aligned}$$

so

$$\Lambda = -\mathbb{E}[W_{a\theta} W_{aa}^{-1} W_{a\theta}] = \mathbb{E}[x^2] =: v_x^2$$

Note that x 's mean could be non-zero. With n dimensions for θ , drawn independently, we have, by the same calculation, $\Lambda = \text{Diag}(\mathbb{E}[x_i^2])$.

Let's now calculate $\kappa[\theta]$. To simplify the notations, we take $\theta_i^d = 0$.

$$\Delta_{\theta_i} = \theta_i W_{\theta_i} = \theta_i x_i (a - \theta \cdot x - \varepsilon)$$

Applying the operator $\Delta_{\eta_a} = \eta_a \partial_a$, $\Delta_{\eta_a} \Delta_{\theta_i} = \eta_a \theta_i x_i$ and

$$\|\Delta_{\theta_i} \Delta_{\eta_a} W\|_\alpha = \|\theta_i x_i \eta_a\|_\alpha = |\theta_i| \|x_i\|_\alpha \|\eta_a\|_\alpha$$

So

$$\kappa[\theta] = \kappa^\theta \sum_i \|\Delta_{\theta_i} \Delta_{\eta_a} W\|_\alpha = \kappa^\theta \sum_i \|\eta_a\|_\alpha |\theta_i| \|x_i\|_\alpha$$

hence $\kappa[\theta] = \sum_i K_i |\theta_i|$ with

$$K_i \equiv \kappa^\theta \|\eta_a\|_\alpha \|x_i\|_\alpha \quad (44)$$

When θ^d is not necessarily 0,

$$\kappa[\theta] = \sum_i K_i |\theta_i - \theta_i^d|$$

So, the maximization (13) is

$$\max_\theta - \sum_i \frac{1}{2} \|x_i\|_2^2 (\theta_i - q_i)^2 - \sum_i K_i |\theta_i - \theta_i^d|$$

We use Lemma 1, which gives (20).

Step 2: The approximate maximization. We calculate $\kappa[a]$ from equation (16). From $W = (a - \theta \cdot x)^2 / 2$, we calculate:

$$\Delta_{\eta_\theta} W = (a - \theta \cdot x) \eta_\theta x, \quad \Delta_a \Delta_{\eta_\theta} W = (a - a^d) \eta_\theta \cdot x$$

hence:

$$\kappa[a] = \kappa^a \|\Delta_a \Delta_{\eta_\theta} W\|_\alpha = \kappa^a |a - a^d| \|\eta_\theta \cdot x\|_\alpha = \kappa^a |a - a^d| \sqrt{\sum_i \sigma_{\theta_i}^2 v_{x_i}^2}$$

All in all, the maximization stage gives:

$$\max_a - \frac{1}{2} (a - \theta \cdot x)^2 - K |a - a^d|$$

with $K = \kappa^a \sqrt{\sum_i \sigma_{\theta_i}^2 v_{x_i}^2}$. This gives

$$a = a^d + \tau (\theta \cdot x - a^d, K) \quad (45)$$

Hence, this example features BR representation of the world, and BR maximization. ■

Proof of Proposition 3. The monopolist solves

$$\max_p \pi(p), \quad \pi(p) = (p - c) (p^d + \tau(p - p^d, \kappa))^{-\psi}$$

Consider first the interior solutions with $p \notin (p^d - \kappa, p^d + \kappa)$. Call $\varepsilon = \text{sign}(p - p^d)$. Then, $p^d + \tau(p - p^d, \kappa) = p - \varepsilon\kappa$ (equation 6). Then, $\partial_p \tau(p - p^d, \kappa) = 1$ and the f.o.c. is $p - \varepsilon\kappa - \psi(p - c) = 0$, i.e.

$$p = p^{int} \equiv \frac{\psi c - \varepsilon \kappa}{\psi - 1} \quad (46)$$

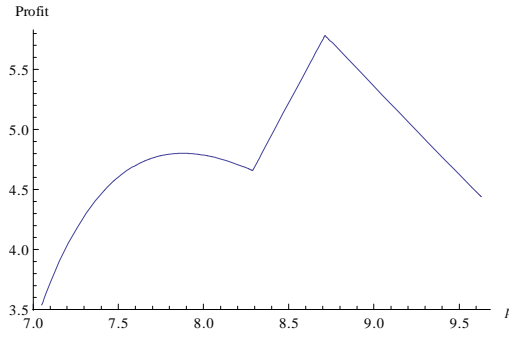


Figure 4: Profit as a function of the price. There are two local maxima: an interior optimum p^{int} and $p = p^d + \kappa > p^{int}$.

The profit is then

$$\pi(p^{int}) = \left(\frac{\psi c - \varepsilon \kappa}{\psi - 1} - c \right) \left(\frac{\psi c - \varepsilon \kappa}{\psi - 1} - \varepsilon \kappa \right)^{-\psi} = \psi^{-\psi} \left(\frac{c - \varepsilon \kappa}{\psi - 1} \right)^{1-\psi}$$

Next, it's not optimal for the monopolist to have $p \notin (p^d - \kappa, p^d + \kappa)$, as $p = p^d + \kappa$ gives the same demand and strictly higher profits. The profit is the

$$\pi(p^d + \kappa) = (p^d + \kappa - c) (p^d)^{-\psi}$$

If it is optimal to choose p^{int} rather than $p^d + \kappa$ iff $R \geq 1$, where

$$\begin{aligned} R(c, c^d, \kappa) &= \frac{\pi(p^{int})}{\pi(p^d + \kappa)} = \frac{\psi^{-\psi} \left(\frac{c - \varepsilon \kappa}{\psi - 1} \right)^{1-\psi}}{\left(\frac{\psi}{\psi - 1} c^d + \kappa - c \right) \left(\frac{\psi}{\psi - 1} c^d \right)^{-\psi}} \\ &= \frac{(c - \varepsilon \kappa)^{1-\psi}}{[\psi c^d + (\psi - 1)(\kappa - c)] (c^d)^{-\psi}} \end{aligned}$$

The c_2 bound is easy: because it is clear (as the profit function is increasing for $p < p^{int}$) that c_2 is such that $p^{int}(c_2) = p^d + \kappa$, i.e. $\frac{\psi c_2 - \kappa}{\psi - 1} = \frac{\psi c^d}{\psi - 1} + \kappa$, i.e. $c_2 = c^d + \kappa$. The trickier case is the case where $c < c^d$, in which case there can be two local maxima. See the illustration below.

Hence, the cutoff c_1 satisfies, with $\varepsilon = -1$,

$$R(c_1, c^d, \kappa) = 1 \tag{47}$$

and $c_1 < c^d$. To obtain an approximate value of c_1 , remark that $R(c, c, 0) = 1$: when $\kappa = 0$, the cutoff corresponds to $c = c^d$. Also, calculations show $\partial_1 R(c, c, \kappa) = 0$ and $\partial_{11} R(c, c, \kappa) \neq 0$. Hence, a small change κ implies a change δc_1 such that, to the leading order, $\frac{1}{2} R_{11} \cdot (\delta c)^2 + R_3 \cdot \kappa = 0$, i.e. $c_1 = c^d - \sqrt{\frac{2R_3 \kappa}{R_{11}}}$. Calculations yield $c_1 = c^d - \sqrt{\frac{2c\kappa}{\psi - 1}} + O(\kappa)$.

References

- Abel, Andrew, Janice C. Eberly and Stavros Panageas. 2010. "Optimal Inattention to the Stock Market with Information Costs and Transactions Costs," Working Paper, Wharton.
- Akerlof, George A., and Janet L. Yellen, "Can Small Deviations from Rationality Make Significant Differences in Economic Equilibria?" *American Economic Review*, 75 (1985), 708-720.
- Bernatzi, Shlomo and Richard Thaler. 2001. "Naive Diversification Strategies in Defined Contribution Saving Plans" *American Economic Review*, Vol. 91, No. 1, 79-98.
- Bernheim, Douglas and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics," *Quarterly Journal of Economics*.
- Brocas, Isabelle and Juan Carillo. 2008. "The Brain as a Hierarchical Organization," *American Economic Review* 2008, 98:4, 1312–1346.
- Camerer, C.F., Ho, T-H, Chong, J-K. "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics*, August 2004, 119(3), 861-898
- Candès, Emmanuel and Terence Tao. "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, 52, 5406-5425.
- Carroll, Gabriel D., James Choi, David Laibson, Brigitte C. Madrian, and Andrew Metrick. 2009. Optimal Defaults and Active Decisions. *Quarterly Journal of Economics*.
- Charness, Gary and Dan Levin, "The Origin of the Winner's Curse: A Laboratory Study," *American Economic Journal: Microeconomics* 2009, 1:1, 207–236
- Chetty, Raj, Adam Looney and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence," *American Economic Review* 99(4): 1145-1177.
- Conlisk, John. 1996. "Why Bounded Rationality?" *Journal of Economic Literature*, 34(2), pp. 669-700
- Crawford, Vincent and Nagore Iriberri. 2007. "Level- k Auctions: Can Boundedly Rational Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?," *Econometrica* 75, 1721–1770
- Dellavigna, Stefano and Ulrike Malmendier. "Paying not to go to the gym " *American Economic Review*, 2006.
- Donoho, David. 2006. "Compressed Sensing", *IEEE Transactions on Information Theory*, 52, 1289-1306.
- Duffie, D., and T. S. Sun. (1990). Transactions costs and portfolio choice in a discrete–continuous-time setting. *Journal of Economic Dynamics & Control* 14(1): 35–51.
- Eyster, Erik, and Matthew Rabin. 2005. "Cursed Equilibrium." *Econometrica*, 73(5):

1623–72.

Eyster, Erik and Georg von Weizacker (2010): “Correlation Neglect in Financial Decision-Making”. Working Paper, LSE.

Fryer, Roland and Matthew O. Jackson (2008) “A Categorical Model of Cognition and Biased Decision Making,” *The B.E. Journal of Theoretical Economics*: Vol. 8: Iss. 1 (Contributions), Article 6

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

Gabaix, Xavier and David Laibson. 2002. “The 6D bias and the Equity Premium Puzzle,” *NBER Macroeconomics Annual*, 16, 257-312.

Gabaix, Xavier, and David Laibson. 2006. “Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets.” *Quarterly Journal of Economics* 121 (2): 505-40.

Gabaix, Xavier, David Laibson, Guillermo Moloche and Stephen Weinberg. 2006. “Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model.” *American Economic Review* 96 (4): 1043-1068.

Gennaioli, Nicola, Andrei Shleifer. 2010. “What Comes to Mind,” *Quarterly Journal of Economics*.

Gennaioli, Nicola, Andrei Shleifer and Robert Vishny. 2010. “Financial Innovation and Financial Fragility,” Working Paper, Harvard.

Huberman, Gur and Wei Jiang. 2006. “Offering versus Choice in 401(k) Plans: Equity Exposure and Number of Funds,” *Journal of Finance*, 61(2), 763–801.

Kahneman, Daniel. “Maps of Bounded Rationality: Psychology for Behavioral Economics,” *American Economic Review*, Vol. 93, No. 5 (Dec., 2003), pp. 1449-1475

Kahneman, Daniel, and Amos Tversky. 1979. “Prospect Theory: An Analysis of Decision under Risk” *Econometrica*, 47, 263-292.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler, "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy*, 97 (1990), 1325-1348.

Kamenica, Emir, “Contextual Inference in Markets: On the Informational Content of Product Lines,” *American Economic Review*, Vol. 98(5), December 2008, pp. 2127–2149.

Kehoe, Patrick and Virgiliu Madrigan. 2010. “Prices Are Sticky After All.,” Working Paper, Minneapolis Fed.

Laibson, David. 1997. “Golden Eggs and Hyperbolic Discounting.” *Quarterly Journal of Economics* 62:443-77.

L’Huillier, Jean-Paul. 2010. Consumers’ Imperfect Information and Nominal Rigidities.

Working paper.

List, John A. “Does Market Experience Eliminate Market Anomalies?,” *Quarterly Journal of Economics*, Vol. 118, No. 1 (Feb., 2003), pp. 41-71

Madrian, Brigitte C., and Dennis Shea, “The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior,” *Quarterly Journal of Economics*, 116 (2001), 1149–1187.

Matejka, Filip. 2010. “Rigid Pricing and Rationally Inattentive Consumer”, Working paper, Princeton.

Mullainathan, Sendhil, “A Memory-Based Model of Bounded Rationality”, *Quarterly Journal of Economics*, 117(3), August 2002: 735-774.

O’Donoghue, Ted and George Loewenstein. 2005. “Animal Spirits: Affective and Deliberative Processes in Economic Behavior,” Working Paper, Cornell.

Palacios-Huerta, I. and O. Volij (2009), “Field Centipedes,” *American Economic Review*, 99(4), 1619-1635.

Reis, Ricardo. 2006. “Inattentive Consumers,” *Journal of Monetary Economics*, 53(6), 1761-1800

Rosenthal, Robert W. 1981. “Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox.” *Journal of Economic Theory*, 25(1): 92–100.

Rubinstein, Ariel. 1998. *Modeling Bounded Rationality*. Cambridge, MA: MIT Press.

Samuelson, W.F., and Max Bazerman. 1985. “The Winner’s Curse in Bilateral Negotiations.” In V. Smith (Ed.), *Research in Experimental Economics*, JAI Press, Volume III, 1985

Samuelson, W.F., and Richard Zeckhauser. 1988, “Status Quo Bias in Decision Making,” *Journal of Risk and Uncertainty*, 1, 7-59

Sargent, Thomas. 1993. *Bounded rationality in macroeconomics*. Oxford University Press.

Shubik, Martin. 1971. “The Dollar Auction Game: A Paradox in Noncooperative Behavior and Escalation,” *The Journal of Conflict Resolution*, Vol. 15, No. 1, pp. 109-111

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

Stahl, Dale O., “Is Step- j Thinking an Arbitrary Modeling Restriction or a Fact of Human Nature?” *Journal of Economic Behavior and Organization*, 37 (1998), 33–51.

Tibshirani, R. (1996). “Regression shrinkage and selection via the lasso.” *J. R. Statist. Soc. B*, 58, 267–288.

Tversky, Amos and Daniel Kahneman. 1981. The Framing of Decisions and the Psychology of Choice,” *Science*, 211(4481), 453-458.