Inattention as a Source of Randomized Discrete Adjustment*

Michael Woodford
Columbia University

April 28, 2008

Abstract

[To be added.]

PRELIMINARY AND INCOMPLETE

*I would like to thank Ricardo Caballero, Eduardo Engel, John Leahy, Filip Matejka, Giuseppe Moscarini, and Alex Wolman for helpful discussions; Maxim Pinkovskiy and Luminita Stevens for outstanding research assistance; the NSF for research support through a grant to the NBER; and the Arthur Okun and Kumho Visiting Professorship, Yale University, for providing the time to write this paper.

In a number of contexts, it has proven useful for economists to model agents as choosing randomly among a discrete set of possible choices, even conditional upon the values of the "fundamental" determinants of choice, such as characteristics of the goods that may be chosen. In macroeconomics, it is common in empirical work on models with fixed costs of adjustment (or other non-convexities) — in contexts such as adjustment of prices, adjustment of a firm's capital stock, or adjustment of the size of its workforce — to assume an "adjustment hazard" (i.e., a function giving the probability of adjustment) that varies continuously with the (real-valued) fundamentals, rather than assuming that adjustment occurs if and only if the fundamentals cross a certain threshold as in "Ss" models (*e.g.,* Caballero and Engel, 1993a, 1993b, 1999). A similar type of randomization is commonly assumed in econometric studies of discrete purchases by households (*e.g.,* McFadden, 1981, 2001; Train, 2003); discrete choice models of this kind are widely used in areas including energy, transportation, environmental studies, health, labor, and marketing.

Often, the assumption of randomization is taken to be merely a proxy for unobserved heterogeneity in either the agents or other aspects of the choice situation; it is supposed that the choice is fully determined by fundamentals, but that some of these are not observed (although one may hypothesize a probability distribution for them). In the discrete-choice literature on individual behavior, it is common to use specifications that can be interpreted as consistent with maximization of a utility function with a random term (McFadden, 1981). In the case of macroeconomic models of adjustment, the adjustment hazard function is sometimes introduced without theoretical foundations (as in Caballero and Engel, 1993a, 1993b); but it is sometimes derived from an assumption that the fixed cost of adjusting is drawn randomly each period from a given probability distribution (*e.g.,* Caballero and Engel, 1999; Dotsey, King and Wolman, 1999).

I believe, however, that there is good reason to prefer an alternative interpretation. First of all, allowance for an arbitrary hazard function is a theory with very weak predictions, which in practice are likely to be made sharper by assuming additional structure purely for analytical convenience, which may well be incorrect. Even when one assumes optimization given a random draw of the fixed cost of adjustment, this is a weak theory in the absence of furhter (unmotivated) assumptions, such as the ubiquitous (but rather implausible) assumption that an independent draw of the fixed cost is made each time. Moreover, experiments suggest that individual choice really is random (*e.g.,* Loomes and Sugden, 1995). Indeed, the recent literature on experimental choice behavior is consistent with a much older literature in the branch of experimental psychology known as "psychophysics," which showed that subjects could not reliably make the same judgment about the relative strength of two stimuli when facing the same choice on repeated occasions (*e.g.*, Fechner, 1859). These experimental data are often explained by models which assume a random element in the perception by the subject of an unchanging stimulus (*e.g.*, Thurstone, 1927); but the randomness is clearly a feature of the subject's nervous system rather than of preferences.

Under the alternative interpretation to be explored here, random choice results precisely from the decisionmaker's difficulty in discriminating among different choice situations, a human cognitive limitation extensively documented by the psychophysicists. (Stochastic perception is already a familiar model of random choice in the experimental psychology literature; see, *e.g.*, Nevin, 1981.) Moreover, rather than assuming an arbitrary random element in agents' perceptions of the choice situation — an arbitrariness that would be subject to the same criticism raised above to random adjustment costs — I shall follow Sims (1998, 2003, 2006) in using information theory to motivate a very specific theory of the nature of the imperfect ability to discriminate among alternative choice situations. This yields not only an extremely parsimonious theory — only one additional free parameter is introduced, relative to the standard "Ss" model (infinite precision of perceptions, constant preferences, constant fixed cost) of discrete choice — but it also allows one to derive randomized choice (conditional on fundamentals) as a *conclusion* of the theory rather than something that is directly assumed (by positing a certain degree of randomness of perceptions).

Under Sims' hypothesis of "rational inattention," a decisionmaker is assumed to have imperfect awareness of the state of the world when making decisions, but the partial information possessed is assumed to correspond to that signal (or set of signals) that would be of *most value* to her (given the decision that she faces), among all possible signals that possess *no greater information content*, in the quantitative sense first defined by Shannon (1948) and used extensively by communications engineers. The idea is that the fundamental bottleneck is the limited attention (or information-processing capacity) of the manager herself, not that some aspects of the state of the world are hidden (or costly to learn about) even if one were to take the trouble to pay close attention to them. This is a highly parsimonious theory, as it introduces only a single new parameter, the assumed limit on information flow to the decisionmaker (in "bits" per time unit), or alternatively, the cost per unit of information (the shadow value of use of the manager's attention for other tasks).[1]

Here I discuss the implications of this hypothesis for discrete choice problems. (The examples studied by Sims in the papers cited instead involve choice of a continuous variable, such as the level of consumption each period. Moreover, a discrete choice problem necessarily requires one to depart from the linear-quadratic-Gaussian framework used by Sims, 1998, 2003.) An important basic result (explained in the next section) is that the optimal information structure (given a cost of information flow) necessarily results in random choice. Moreover, under fairly simple assumptions, the precise model of random choice implied is a logit specification often assumed in econometric models of household choice.

1 Rational Inattention and the Optimal Adjustment Hazard

In order to show how "rational inattention" of the sort hypothesized by Sims (1998, 2003, 2006) gives rise to a continuous "adjustment hazard" of the kind postulated by Caballero and Engel (1993a, 1993b), I here consider a simple one-time choice between two alternatives. A leading example of the kind of problem that I wish to study is a firm's choice of the times at which to review its pricing policy. (This application is treated in detail in Woodford, 2008.) In the discussion here, I shall describe the problem as one of choosing whether or not to reconsider the firm's existing price, but the same formalism could be applied to any of many different discrete choice problems.

1.1 Formulation of the Problem

Let the "normalized price" of a firm i be defined as $q(i) \equiv \log(p(i)/PY)$, where $p(i)$ is the price charged by firm i for its product, P is an aggregate price index, and Y is an index of aggregate output (or aggregate real expenditure), and suppose that

[1]Other applications of this approach include Mackowiak and Wiederholt (2007), Paciello (2007), and Tutino (2007). Other recent models of endogenous choice of how well informed decisionmakers should be include Moscarini (2004), Bacchetta and Van Wincoop (2005), and Reis (2006a, 2006b).

the expected payoff[2] to the firm of charging normalized price q is given by a function $V(q)$, which achieves its maximum value at the optimal normalized price

$$q^* \equiv \arg\max_q V(q).$$

I shall assume that $V(q)$ is a smooth, strictly quasi-concave function. By *strict* quasi-concavity, I mean that not only are the sets $\{q|V(q) \geq v\}$ convex for all v, but in addition the sets $\{q|V(q) = v\}$ are of (Lebesgue) measure zero. Strict quasi-concavity implies that there exists a smooth, monotonic transformation $q = \phi(\hat{q})$ such that the function $\hat{V}(\hat{q}) \equiv V(\phi(\hat{q}))$ is not only a concave function, but a *strictly* concave function of $\hat{q}$. In this case, under the further assumption that $V(q)$ achieves a maximum, the maximum q^* must be unique. Moreover, q^* is the unique point at which $V'(q^*) = 0$; and one must have $V'(q) > 0$ for all $q < q^*$, while $V'(q) < 0$ for all $q > q^*$.

We can then define a "price gap" $x(i) \equiv q(i) - q^*$, as in Caballero and Engel, indicating the signed discrepancy between a firm's actual price and the price that it would be optimal for it to charge.[3] Under full information and in the absence of any cost of changing its price, a firm should choose to set $q(i) = q^*$. Let us suppose, though, that the firm must pay a fixed cost $\kappa > 0$ in order to conduct a review of its pricing policy. I shall suppose, as in canonical menu-cost models, that a firm that conducts such a review learns the precise value of the current optimal price, and therefore adjusts its price so that $q(i) = q^*$. A firm that chooses not to review its existing policy instead continues to charge the price that it chose on the occasion of its last review of its pricing policy. The loss from failing to review the policy (or alternatively, the gain from reviewing it, net of the fixed cost) is then given by

$$L(x) \equiv V(q^*) - V(q^* + x) - \kappa, \tag{1.1}$$

[2] I need not be specific at this stage about the nature of this payoff. In the eventual dynamic problem considered below, it includes not only profits in the current period (when the price $p(i)$ is charged), but also the implications for expected discounted profits in later periods of having chosen a price $p(i)$ in the current period.

[3] It might appear simpler to directly define the normalized price as the price relative to the optimal price, rather than relative to aggregate nominal expenditure, so that the optimal normalized price would be zero, by definition. But the optimal value q^* is something that we need to determine, rather than something that we know at the time of introducing our notation. (Eventually, the function $V(q)$ must be endogenously determined, as discussed in section 2 below.)

as a function of the price gap x that exists prior to the review.

If $V(q)$ is a smooth, strictly quasi-concave function, then $L(x)$ is a smooth, strictly quasi-convex function, with a unique minimum at $x = 0$. Then in the case of full information, the optimal price-review policy is to review the price if and only if the value of x prior to the review is in the range such that $L(x) \geq 0$.[4] The values of x such that a price review occurs will consist of all x *outside* a certain interval, the "zone of inaction," which necessarily includes a neighborhood of the point $x = 0$. The boundaries of this interval (one negative and one positive, in the case that the interval is bounded) constitute the two "Ss triggers" of an "Ss model" of price adjustment.

I wish now to consider instead the case in which the firm does *not* know the value of x prior to conducting the review of its policy. I shall suppose that the firm *does* know its existing price, so that it is possible for it to continue to charge that price in the absence of a review; but it does not know the current value of aggregate nominal expenditure PY, and so does not know its *normalized* price, or the gap between its existing price and the currently optimal price. I shall furthermore allow the firm to have *partial* information about the current value of x prior to conducting a review; this is what I wish to motivate as optimal subject to limits on the attention that the firm can afford to pay to market conditions between the occasions when the fixed cost κ is paid for a full review. It is on the basis of this partial information that the decision whether to conduct a review must be made.

Following Sims, I shall suppose that absolutely *any* information about current (or past) market conditions can be available to the firm, as long as the quantity of information obtained by the firm outside of a full review is within a certain finite limit, representing the scarcity of attention, or information-processing capacity, that is deployed for this purpose. The quantity of information obtained by the firm in a given period is defined as in the information theory of Claude Shannon (1948), used extensively by communications engineers. In this theory, the quantity of information contained in a given signal is measured by the reduction in entropy of the decisionmaker's posterior over the state space, relative to the prior distribution. Let us suppose that we are interested simply in information about the current value

[4]The way in which we break ties in the case that $L(x) = 0$ exactly is arbitrary; here I suppose that in the case of indifference the firm reviews its price. In the equilibrium eventually characterized below for the full-information case, values of x for which $L(x) = 0$ exactly occur with probability zero, so this arbitrary choice is of no consequence.

of the unknown (random) state x, and that the firm's *prior* is given by a density function $f(x)$ defined on the real line.[5] Let $\hat{f}(x|s)$ instead be the firm's posterior, conditional upon observing a particular signal s. The *entropy* associated with a given density function (a measure of the degree of uncertainty with a number of attractive properties) is equal to[6]

$$-\int f(x)\log f(x)dx,$$

and as a consequence the entropy reduction when signal s is received is given by

$$I(s) \equiv \int \hat{f}(x|s)\log \hat{f}(x|s)dx - \int f(x)\log f(x)dx.$$

The average information revealed by this kind of signal is therefore

$$I \equiv \mathrm{E}_s I(s) \qquad (1.2)$$

where the expected value is taken over the set of possible signals that were possible *ex ante*, using the prior probabilities of that each of these signals would be observed.[7] It is this total quantity I that determines the bandwidth (in the case of radio signals, for example), or the *channel capacity* more generally (an engineering limit of any communication system), that must be allocated to the transmission of this signal if the transmission of a signal with a given average information content is to be possible.[8] Sims correspondingly proposes that the limited attention of decisionmakers

[5] In section 1.2, we consider what this prior should be, if the firm understands the process that generates the value of x, but has not yet obtained any information about current conditions. For now, the prior is arbitrarily specified as some pre-existing state of knowledge that does not precisely identify the state x.

[6] In information theory, it is conventional to define entropy using logarithms with a base of two, so that the quantity I defined in (1.2) measures information in "bits", or binary digits. (One bit is the amount of information that can be transmitted by the answer to one optimally chosen yes/no question, or by revealing whether a single binary digit is 0 or 1.) I shall instead interpret the logarithm in this and subsequent formulas as a natural logarithm, to allow the elimination of a constant in various expressions. This is an equivalent measure of information, but with a different size of unit: one unit of information under the measure used here (sometimes called a "nat") is equivalent to 1.44 bits of information.

[7] The prior over s is the one implied by the decisionmaker's prior over possible values of x, together with the known statistical relationship between the state x and the signal s that will be received.

[8] Shannon's theorems pertain to the relation between the properties of a given communication channel and the *average* rate at which information can be transmitted over time using that channel, not the amount of information that will be contained in the signal that is sent over any given short time interval.

be modeled by assuming a constraint on the possible size of the average information flow I.

I shall suppose, then, that the firm arranges to observe a signal s before deciding whether to pay the cost κ and conduct a review of its pricing policy. The theory of *rational inattention* posits that both the design of this signal (the set of possible values of s, and the probability that each will be observed conditional upon any given state x) and the decision about whether to conduct a price review conditional upon the signal observed will be optimal, in the sense of maximizing

$$\bar{L} \equiv \mathrm{E}[\delta(s)L(x)] - \theta I, \tag{1.3}$$

where $\delta(s)$ is a (possibly random) function of s indicating whether a price review is undertaken ($\delta = 1$ when a price review occurs, and $\delta = 0$ otherwise); the expectation operator integrates over possible states x, possible signals s, and possible price-review decisions, under the firm's prior; and $\theta > 0$ is a cost per unit of information of being more informed when making the price-review decision. (This design problem is solved from an *ex ante* perspective: one must decide how to allocate one's attention, which determines what kind of signal one will observe under various circumstances, before learning anything about the current state.)

I have here written the problem as if a firm can allocate an arbitrary amount of attention to tracking market conditions between full price reviews, and hence have an estimate of x of arbitrary precision prior to its decision about whether to conduct the review, if it is willing to pay for this superior information. One might alternatively consider the problem of choosing a partial information structure to maximize $\mathrm{E}[\delta(s)L(x)]$ subject to an upper bound on I. This will lead to exactly the same one-parameter family of informationally-efficient policies, indexed by the value of I rather than by the value of θ. (In the problem with an upper bound on the information used, there will be a unique value of θ associated with each informationally-efficient policy, corresponding to the Lagrange multiplier for the constraint on the value of I; there will be an inverse one-to-one relationship between the value of θ and the value of I.) I prefer to consider the version of the problem in which θ rather than I is given as part of the specification of the environment. This is because decisionmakers have much more attention to allocate than the attention allocated to any one task, and could certainly allocate more attention to aspects of market conditions relevant to the scheduling of reviews of pricing policy, were this of sufficient importance; it makes more sense to suppose that there is a given shadow price of additional attention, determined by the opportunity cost of reducing the attention paid to other matters, rather than a fixed bound on the attention that can be paid to the problem considered here, even if there *is* a global bound on the information-processing capacity of the decisionmaker.

1.2 Characterization of the Solution

I turn now to the solution of this problem, taking as given the prior $f(x)$, the loss function $L(x)$, and the information cost $\theta > 0$. A first observation is that an efficient signal will supply no information other than whether the firm should review its pricing policy.

Lemma 1 *Consider any signalling mechanism, described by a set of possible signals* S *and conditional probabilities* $\pi(s|x)$ *for each of the possible signals* $s \in S$ *in each of the possible states* x *in the support of the prior* f*, and any decision rule, indicating for each* $s \in S$ *the probability* $p(s)$ *with which a review occurs when signal* s *is observed. Let* $\bar{L}$ *be the value of the objective (1.3) implied by this policy on the part of the firm. Consider as well the alternative policy, under which the set of possible signals is* $\{0, 1\}$*, the conditional probability of receiving the signal 1 is*

$$\pi(1|x) = \int_{s \in S} p(s)\pi(s|x)ds$$

for each state x *in the support of* f*, and the decision rule is to conduct a review with probability one if and only if the signal 1 is observed; and let* $\bar{L}^*$ *be the value of (1.3) implied by this alternative policy. Then* $\bar{L}^* \geq \bar{L}$*.*

Moreover, the inequality is strict, except if the first policy is one under which either (i) $\pi(s|x)$ *is independent of* x *(almost surely), so that the signals convey no information about the state* x*; or (ii)* $p(s)$ *is equal to either zero or one for all signals that occur with positive probability, and the conditional probabilities are of the form*

$$\pi(s|x) = \pi(s|p(s)) \cdot \pi(p(s)|x),$$

where the conditional probability $\pi(s|p(s))$ *of a given signal* s *being received, given that the signal will be one of those for which* $p(s)$ *takes a certain value, is independent of* x

(almost surely). That is, either the original signals are completely uninformative; or the original decision rule is deterministic (so that the signal includes a definite recommendation as to whether a price review should be undertaken) and any additional information contained in the signal, besides the implied recommendation regarding the price-review decision, is completely uninformative.

A proof is given in Appendix A. Note that this result implies that we may assume, without loss of generality, that an optimal policy involves only two possible signals, $\{0, 1\}$, and a decision rule under which a review is scheduled if and only if the signal 1 is received. That is, the only signal received is an indication whether it is time to review the firm's existing price or not. (If the firm arranges to receive any more information than this, it is wasting its scarce information-processing capacity.) A policy of this form is completely described by specifying the *hazard function* $\Lambda(x) \equiv \pi(1|x)$, indicating the probability that a price review occurs, in the case of any underlying state x in the support of f.

It follows from Lemma 1 that any randomization that is desired in the price-review decision should be achieved by arranging for the *signal* about market conditions to be random, rather than through any randomization by the firm after receiving the signal. This does not, however, imply in itself that the signal that determines the timing of price reviews should be random, as in the Calvo model (or the "generalized Ss model" of Caballero and Engel). But in fact one can show that it *is* optimal for the signal to be random, under extremely weak conditions.

Let us consider the problem of choosing a measurable function $\Lambda(x)$, taking values on the interval $[0, 1]$, so as to maximize (1.3). One must first be able to evaluate (1.3) in the case of a given hazard function. This is trivial when $\Lambda(x)$ is (almost surely) equal to either 0 or 1 for all x, as in either case the information content of the signal is zero. Hence $\bar{L} = E[L(x)]$ if $\Lambda(x) = 1$ (a.s), and $\bar{L} = 0$ if $\Lambda(x) = 0$ (a.s.). After disposing of these trivial cases, we turn to the case in which the prior probability of a price review

$$\bar{\Lambda} \equiv \int \Lambda(x) f(x) dx \tag{1.4}$$

takes an interior value, $0 < \bar{\Lambda} < 1$. As there are only two possible signals, there are two possible posteriors, given by

$$\hat{f}(x|0) = \frac{f(x)(1 - \Lambda(x))}{1 - \bar{\Lambda}}, \qquad \hat{f}(x|1) = \frac{f(x)\Lambda(x)}{\bar{\Lambda}}$$

using Bayes' Law. The information measure I is then equal to

$$\begin{aligned} I &= \bar{\Lambda} I(1) + (1-\bar{\Lambda}) I(0) \\ &= \bar{\Lambda} \int \hat{f}(x|1) \log \hat{f}(x|1) dx + (1-\bar{\Lambda}) \int \hat{f}(x|0) \log \hat{f}(x|0) dx - \int f(x) \log f(x) dx \\ &= \int \varphi(\Lambda(x)) f(x) dx - \varphi(\bar{\Lambda}), \end{aligned} \tag{1.5}$$

where

$$\varphi(\Lambda) \equiv \Lambda \log \Lambda + (1-\Lambda) \log(1-\Lambda) \tag{1.6}$$

in the case of any $0 < \Lambda < 1$, and we furthermore define[9]

$$\varphi(0) = \varphi(1) = 0.$$

We can therefore rewrite the objective (1.3) in this case as

$$\bar{L} = \int [L(x)\Lambda(x) - \theta\varphi(\Lambda(x))] f(x) dx + \theta\varphi(\int \Lambda(x) f(x) dx). \tag{1.7}$$

Given the observation above about the trivial cases, the same formula applies as well when $\bar{\Lambda}$ is equal to 0 or 1. Hence (1.7) applies in the case of any measurable function $\Lambda(x)$ taking values in $[0, 1]$, and our problem reduces to the choice of $\Lambda(x)$ to maximize (1.7).

This is a problem in the calculus of variations. Suppose that we start with a function $\Lambda(x)$ such that $0 < \bar{\Lambda} < 1$, and let us consider the effects of an infinitesimal variation in this function, replacing $\Lambda(x)$ by $\Lambda(x) + \delta\Lambda(x)$, where $\delta\Lambda(x)$ is a bounded, measurable function indicating the variation. We observe that

$$\delta\bar{L} = \int \partial(x) \cdot \delta\Lambda(x)\, f(x) dx$$

where

$$\partial(x) \equiv L(x) - \theta\varphi'(\Lambda(x)) + \theta\varphi'(\bar{\Lambda}).$$

[9] This definition follows Shannon (1948); our $\varphi(\Lambda)$ is the negative of his "binary entropy function." Note that under this extension of the definition of $\varphi(\Lambda)$ to the boundaries of its domain, the function is continuous on the entire interval. Moreover, under this definition, (1.5) is a correct measure of the information content of the signal (namely, zero) even in the case that one of the signals occurs with probability zero.

A first-order condition for (local) optimality of the policy is then at each point x (almost surely[10]), one of the following conditions holds: either $\Lambda(x) = 0$ and $\partial(x) \leq 0$; $\Lambda(x) = 1$ and $\partial(x) \geq 0$; or $0 < \Lambda(x) < 1$ and $\partial(x) = 0$. We can furthermore observe from the behavior of the function $\varphi'(\Lambda) = \log(\Lambda/1 - \Lambda)$ near the boundaries of the domain that

$$\lim_{\Lambda(x)\to 0} \partial(x) = +\infty, \qquad \lim_{\Lambda(x)\to 1} \partial(x) = -\infty,$$

so that neither of the first two conditions can ever hold. Hence the first-order condition requires that

$$\partial(x) = 0 \tag{1.8}$$

almost surely.

This condition implies that

$$\frac{\Lambda(x)}{1 - \Lambda(x)} = \frac{\bar{\Lambda}}{1 - \bar{\Lambda}} \exp\left\{\frac{L(x)}{\theta}\right\} \tag{1.9}$$

for each x. Condition (1.9) implicitly defines a measurable function $\Lambda(x) = \Lambda^*(x; \bar{\Lambda})$ taking values in $(0, 1)$.[11] It is worth noting that in this solution, for a fixed value of $\bar{\Lambda}$, $\Lambda(x)$ is monotonically increasing in the value of $L(x)/\theta$, approaching the value 0 for large enough negative values of $L(x)/\theta$, and the value 1 for large enough positive values; and for given x, $\Lambda^*(x; \bar{\Lambda})$ is an increasing function of $\bar{\Lambda}$, approaching 0 for values of $\bar{\Lambda}$ close enough to 0, and 1 for values of $\bar{\Lambda}$ close enough to 1. We can extend the definition of this function to extreme values of $\bar{\Lambda}$ by defining

$$\Lambda^*(x; 0) = 0, \qquad \Lambda^*(x; 1) = 1$$

for all values of x; when we do so, $\Lambda^*(x; \bar{\Lambda})$ remains a function that is continuous in both arguments.

The above calculation implies that in the case of any (locally) optimal policy for which $0 < \bar{\Lambda} < 1$, the hazard function must be equal (almost surely) to a member of the one-parameter family of functions $\Lambda^*(x; \bar{\Lambda})$. It is also evident (from definition (1.4) and the bounds that $\Lambda(x)$ must satisfy) that if $\bar{\Lambda}$ takes either of the extreme

[10] Note that we can only expect to determine the optimal hazard function $\Lambda(x)$ up to arbitrary changes on a set of values of x that occur with probability zero under the prior, as such changes have no effect on any of the terms in the objective (1.7).

[11] We can easily give a closed-form solution for this function: $\Lambda^*(x; \bar{\Lambda}) = R/1 + R$, where R is the right-hand side of (1.9).

values 0 or 1, the hazard function must satisfy $\Lambda(x) = \bar{\Lambda}$ almost surely; hence the hazard function would be equal (almost surely) to a member of the one-parameter family in these cases as well. We can therefore conclude that the optimal hazard function must belong to this family; it remains only to determine the optimal value of $\bar{\Lambda}$.

In this discussion, $\bar{\Lambda}$ has been used both to refer to the value defined in (1.4) and to index the members of the family of hazard functions defined by (1.9). In fact, the same numerical value of $\bar{\Lambda}$ must be both things. Hence we must have

$$J(\bar{\Lambda}) = \bar{\Lambda}, \tag{1.10}$$

where

$$J(\bar{\Lambda}) \equiv \int \Lambda^*(x;\bar{\Lambda}) f(x) dx. \tag{1.11}$$

Condition (1.10) necessarily holds in the case of a locally optimal policy, but it does not guarantee that $\Lambda^*(x;\bar{\Lambda})$ is even locally optimal. We observe from the definition that $J(0) = 0$ and $J(1) = 1$, so $\bar{\Lambda} = 0$ and $\bar{\Lambda} = 1$ are always at least two solutions to equation (1.10); yet these need not be even local optima.

We can see this by considering the function $\bar{L}(\bar{\Lambda})$, obtained by substituting the solution $\Lambda^*(x;\bar{\Lambda})$ defined by (1.9) into the definition (1.7). Since any locally optimal policy must belong to this one-parameter family, an optimal policy corresponds to a value of $\bar{\Lambda}$ that maximizes $\bar{L}(\bar{\Lambda})$. Differentiating this function, we obtain

$$\begin{aligned} \bar{L}'(\bar{\Lambda}) &= \int [L(x) - \theta\varphi'(\Lambda^*(x))]\Lambda^*_{\bar{\Lambda}}(x) f(x) dx + \theta\varphi'(J(\bar{\Lambda})) \int \Lambda^*_{\bar{\Lambda}}(x) f(x) dx \\ &= \theta[\varphi'(J(\bar{\Lambda})) - \varphi'(\bar{\Lambda})] \int \Lambda^*_{\bar{\Lambda}}(x) f(x) dx, \end{aligned}$$

at any point $0 < \bar{\Lambda} < 1$, where $\Lambda^*_{\bar{\Lambda}}(x) > 0$ denotes the partial derivative of $\Lambda^*(x;\bar{\Lambda})$ with respect to $\bar{\Lambda}$, and we have used the first-order condition $\partial(x) = 0$, satisfied by any hazard function in the family defined by (1.9), to obtain the second line from the first. Since

$$\int \Lambda^*_{\bar{\Lambda}}(x) f(x) dx > 0,$$

it follows that $\bar{L}'(\bar{\Lambda})$ has the same sign as $\varphi'(J(\bar{\Lambda})) - \varphi'(\bar{\Lambda})$, which (because of the monotonicity of $\varphi'(\Lambda)$), has the same sign as $J(\bar{\Lambda}) - \bar{\Lambda}$.

Hence a value of $\bar{\Lambda}$ that satisfies (1.10) corresponds to a critical point of $\bar{L}(\bar{\Lambda})$, but not necessarily to a local maximum. The complete set of necessary and sufficient conditions for a local maximum are instead that $\Lambda(x)$ be a member of the one-parameter family of hazard functions defined by (1.9), for a value of $\bar{\Lambda}$ satisfying (1.10), and such that in addition, (i) if $\bar{\Lambda} > 0$, then $J(\Lambda) > \Lambda$ for all Λ in a left neighborhood of $\bar{\Lambda}$; and (ii) if $\bar{\Lambda} < 1$, then $J(\Lambda) < \Lambda$ for all Λ in a right neighborhood of $\bar{\Lambda}$. The argument just given only implies that there must exist solutions with this property, and that they correspond to at least locally optimal policies. In fact, however, there is necessarily a unique solution of this form, and it corresponds to the global optimum, owing to the following result.

Lemma 2 *Let the loss function $L(x)$, the prior $f(x)$, and the information cost $\theta > 0$ be given, and suppose that $L(x)$ is not equal to zero almost surely [under the measure defined by f].*[12] *Then the function $J(\Lambda)$ has a graph of one of three possible kinds: (i) if*

$$\int \exp\left\{\frac{L(x)}{\theta}\right\} f(x)dx \leq 1, \qquad \int \exp\left\{-\frac{L(x)}{\theta}\right\} f(x)dx > 1,$$

then $J(\Lambda) < \Lambda$ for all $0 < \Lambda < 1$ [as in the first panel of Figure 1], and the optimal policy corresponds to $\bar{\Lambda} = 0$; (ii) if

$$\int \exp\left\{-\frac{L(x)}{\theta}\right\} f(x)dx \leq 1, \qquad \int \exp\left\{\frac{L(x)}{\theta}\right\} f(x)dx > 1,$$

then $J(\Lambda) > \Lambda$ for all $0 < \Lambda < 1$ [as in the second panel of Figure 1], and the optimal policy corresponds to $\bar{\Lambda} = 1$; and (iii) if

$$\int \exp\left\{\frac{L(x)}{\theta}\right\} f(x)dx > 1, \qquad \int \exp\left\{-\frac{L(x)}{\theta}\right\} f(x)dx > 1,$$

then there exists a unique interior value $0 < \bar{\bar{\Lambda}} < 1$ at which $J(\bar{\bar{\Lambda}}) = \bar{\bar{\Lambda}}$, while $J(\Lambda) > \Lambda$ for all $0 < \Lambda < \bar{\bar{\Lambda}}$, and $J(\Lambda) < \Lambda$ for all $\bar{\bar{\Lambda}} < \Lambda < 1$ [as in the third panel of Figure 1], and the optimal policy corresponds to $\bar{\Lambda} = \bar{\bar{\Lambda}}$.

The proof is again in Appendix A. Note that the three cases considered in the lemma exhaust all possibilities, as it is not possible for both of the integrals to simultaneously

[12] This is a very weak assumption. Note that it would be required by the assumption invoked earlier, that $L(x)$ is strictly quasi-concave. But in fact, since $L(0) = -\kappa$, it suffices that the loss function be continuous at zero and that $f(x)$ be positive on a neighborhood of zero, though even these conditions are not necessary.

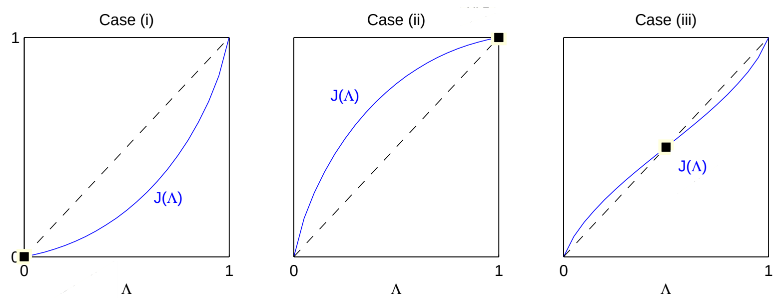

Figure 1: The three possible shapes of the function $J(\Lambda)$, as explained in Lemma 2. In each case, the optimal value of $\bar{\Lambda}$ is indicated by the black square.

have a value no greater than 1 (in the case that $L(x)$ is not equal to zero almost surely), as a consequence of Jensen's Inequality. Thus we have given a complete characterization of the optimal policy.

Our results also provide a straightforward approach to computation of the optimal policy, once the loss function $L(x)$, the prior $f(x)$, and the value of θ are given. Given $L(x)$ and θ, (1.9) allows us to compute $\Lambda^*(x; \bar{\Lambda})$ for any value of $\bar{\Lambda}$; given $f(x)$, it is then straightforward to evaluate $J(\Lambda)$ for any $0 < \Lambda < 1$, using (1.11). Finally, once one plots the function $J(\Lambda)$, it is easy to determine the optimal value $\bar{\Lambda}$; Lemma 2 guarantees that a simple bisection algorithm will necessarily converge to the right fixed point, as discussed in Appendix B.

1.3 Discussion

We can now see that the optimal signalling mechanism necessarily involves randomization, as remarked earlier. In any case in which it is optimal neither to *always* review one's price nor to *never* review one's price, so that the average frequency with which price reviews occur is some $0 < \bar{\Lambda} < 1$, the optimal hazard function satisfies $0 < \Lambda(x) < 1$, so that a price review may or may not occur, in the case of *any* current price gap x.[13] This is not simply an assumption. We have allowed for the possibility of a hazard function which takes the value 0 on some interval (the "zone of inaction") in which x falls with a probability $1 - \bar{\Lambda}$, and the value 1 everywhere outside that interval; but this can never be an optimal policy. Hence an optimal signalling mechanism never provides a signal that is a *deterministic* function of the true state.

[13] As usual, the qualification "almost surely" must be added.

One can also easily show that our assumption that the signal must be a random function of the current state x alone; that is, the randomness in the relation between the observed signal and the value of x must be purely uninformative about the state of the world — it must represent noise in the measurement process itself, rather than systematic dependence on some other aspect of the current (or past) state of the world. We could easily consider a mechanism in which the probability of receiving a given signal s may depend on both x and some other state y. (Statement of the problem then requires that the prior $f(x, y)$ over the joint distribution of the two states be specified.) The same argument as above implies that an optimal policy can be described by a hazard function $\Lambda(x, y)$, and that the optimal hazard function will again be of the form (1.9), where one simply replaces the argument x by (x, y) everywhere. In the case that the value function depends only on the state x, as assumed above, the loss function will also be a function simply of x; hence (1.9) implies that the optimal hazard will depend only on x, and that it will be the function of x characterized above.

Among the consequences of this result is the fact that the random signals received by different firms, each of which has the same prior $f(x)$ about its current price gap, will be distributed *independently* of one another, as assumed in the Calvo model. If the signals received by firms were instead correlated (for example, if with probability $\bar{\Lambda}$ all firms receive a signal to review their prices, while with probability $1 - \bar{\Lambda}$ none of them do), then each firm's signal would convey information about *other firms' signals,* and also about their actions. Such signals would therefore convey more information (and, under our assumption about the cost of information, necessarily be more costly) than uncorrelated signals, without being any more useful to the firms in helping them to make profit-maximizing decisions; the correlated signals would therefore not represent an efficient signalling mechanism.[14] Hence the present model predicts that while the price-review decision is random at the level of an individual firm, the *fraction* of such firms that will review their prices in aggregate (assuming a large enough number of firms for the law of large numbers to apply) will be $\bar{\Lambda}$ with

[14]Of course, this result depends on an assumption that, as in the setup assumed by Caballero and Engel (1993a, 2007), the payoff to a firm depends only on its *own* normalized price, and not also on the relation between its price and the prices of other imperfectly attentive firms; to the extent that information about others' actions is payoff-relevant, an optimal signalling mechanism *will* involve correlation.

certainty.

The present model provides a decision-theoretic justification for the kind of "generalized Ss" behavior proposed by Caballero and Engel (1993a, 1993b) as an empirical specification. The interpretation is different from the hypothesis of a random menu cost in Caballero and Engel (1999) and Dotsey, King and Wolman (1999), but the present model is observationally equivalent to a random-menu-cost model, in the case that the distribution of menu costs belongs to a particular one-parameter family. Suppose that firm has perfect information, but that the menu cost $\tilde{\kappa}$ is drawn from a distribution with cumulative distribution function $G(\tilde{\kappa})$, rather than taking a certain positive value κ with certainty. Then a firm with price gap x should choose to revise its price if and only if

$$V(q^*) - \tilde{\kappa} \geq V(q^* + x),$$

which occurs with probability

$$\Lambda(x) = G(V(q^*) - V(q^* + x)) = G(L(x) + \kappa), \tag{1.12}$$

where once again $L(x)$ is the loss function (1.1) of a firm with constant menu cost κ. Thus (1.12) is the hazard function implied by a random-menu-cost model; the only restriction implied by the theory is that $\Lambda(x)$ must be a non-decreasing function of the loss $L(x)$. The present theory also implies that $\Lambda(x)$ should be a non-decreasing function of $L(x)$, as (1.9) has this property for each value of $\bar{\Lambda}$. In fact, the optimal hazard function under rational inattention is identical to the hazard function of a random-menu-cost model in which the distribution of possible menu costs is given by[15]

$$G(\tilde{\kappa}) = 1 - \left[1 + \left(\frac{\bar{\Lambda}}{1 - \bar{\Lambda}}\right) \exp\left\{\frac{\tilde{\kappa} - \kappa}{\theta}\right\}\right]^{-1}. \tag{1.13}$$

While the present model does not imply behavior inconsistent with a random-menu-cost model, it makes much sharper predictions. Moreover, not only does the present model correspond to a single very specific one-parameter family of possible distributions of menu costs, but these distributions are all fairly different from what is usually assumed in calibrations of random-menu-cost models. In particular, a distribution of the form (1.13) necessarily has an atom at zero, so that the hazard is bounded away from zero even for values of x near zero; it has instead been common

[15]Note that in this formula, κ is a parameter of the distribution, not the size of the menu cost.

in numerical analyses of generalized Ss models to assume that in a realistic specification there should be no atom at zero, so that $\Lambda(0) = 0$. The fact that the present model instead implies that $\Lambda(0)$ is necessarily positive (if price reviews occur with any positive frequency) — and indeed, may be a substantial fraction of the average frequency $\bar{\Lambda}$ — is an important difference; under the rule of thumb discussed by Caballero and Engel (2006), it reduces the importance of the "extensive margin" of price adjustment, and hence makes the predictions of a generalized Ss model more similar to those of the Calvo model.

The random-menu-cost model also provides no good reason why, in a dynamic extension of the model, the adjustment hazard should depend only on the current price gap x, and not also on the time elapsed since the last price review. This case is *possible*, of course, if one assumes that the menu cost $\tilde{\kappa}$ is drawn *independently* each period from the distribution G. But there is no reason to assume such independence, and the specification does not seem an especially realistic one (though obviously convenient from the point of view of empirical tractability), if the model is genuinely about exogenous time variation in the cost of changing one's price. The theory of rational inattention instead *requires* that the hazard rate depend only on the current state x, as long as the dynamic decision problem is one in which both the prior and the value function are stationary over time (rather than being duration-dependent), as in the dynamic model developed in the next section.

2 Microfoundations for the Calvo Model of Price Adjustment

[To be added.]

References

[1] Bacchetta, Philippe, and Eric Van Wincoop, "Rational Inattention: A Solution to the Forward Discount Puzzle," NBER working paper no. 11633, September 2005.

[2] Bonomo, Marco, and Carlos Carvalho, "Endogenous Time-Dependent Rules and Inflation Inertia," *Journal of Money, Credit and Banking* 36: 1015-1041 (2004).

[3] Burstein, Ariel T., and Christian Hellwig, "Prices and Market Share in a Menu Cost Model," unpublished, U.C.L.A., September 2007.

[4] Caballero, Ricardo J., "Time Dependent Rules, Aggregate Stickiness, and Information Externalities," Discussion Paper no. 428, Department of Economics, Columbia University, May 1989.

[5] Caballero, Ricardo J., and Eduardo Engel, "Microeconomic Rigidities and Aggregate Price Dynamics," *European Economic Review* 37: 697-717 (1993a).

[6] Caballero, Ricardo J., and Eduardo Engel, "Microeconomic Adjustment Hazards and Aggregate Dynamics," *Quarterly Journal of Economics* 108: 359-383 (1993b).

[7] Caballero, Ricardo J., and Eduardo Engel, "Explaining Investment Dynamics in U.S. Manufacturing: A Generalized *(S,s)* Approach," *Econometrica* 67: 783-826 (1999).

[8] Caballero, Ricardo J., and Eduardo Engel, "Price Stickiness in Ss Models: Basic Properties," unpublished, M.I.T., October 2006.

[9] Caballero, Ricardo J., and Eduardo Engel, "Price Stickiness in Ss Models: New Interpretations of Old Results," *Journal of Monetary Economics* 54 (supp. 1): 100-121 (2007).

[10] Calvo, Guillermo A., "Staggered Prices in a Utility Maximizing Framework," *Journal of Monetary Economics* 12: 383-398 (1983).

[11] Caplin, Andrew S., and Daniel F. Spulber, "Menu Costs and the Neutrality of Money," *Quarterly Journal of Economics* 102: 703-726 (1987).

[12] Cover, Thomas M., and Joy A. Thomas, *Elements of Information Theory,* New York: Wiley-Interscience, 2d ed., 2006.

[13] Dotsey, Michael and Robert G. King, "Implications of State-Dependent Pricing for Dynamic Macroeconomic Models," *Journal of Monetary Economics* 52: 213-242 (2005).

[14] Dotsey, Michael, Robert G. King, and Alexander L. Wolman, "State Dependent Pricing and the General Equilibrium Dynamics of Money and Output," *Quarterly Journal of Economics* 114: 655-690 (1999).

[15] Fechner, Gustav T., *Elemente der Psychophysik,* Leipzig: Breitkopf and Hartel, 1859.

[16] Gertler, Mark, and John Leahy, "A Phillips Curve with an Ss Foundation," unpublished, New York University, May 2007.

[17] Golosov, Michael, and Robert E. Lucas, Jr., "Menu Costs and Phillips Curves," *Journal of Political Economy* 115: 171-199 (2007).

[18] Kiley, Michael T., "Partial Adjustment and Staggered Price Setting," *Journal of Money, Credit and Banking* 34: 283-298 (2002).

[19] Levin, Andrew, and Tack Yun, "Reconsidering the Natural Rate Hypothesis in a New Keynesian Framework," *Journal of Monetary Economics* 54: 1344-1365 (2007).

[20] Loomes, Graham, and Robert Sugden, "Incorporating a Stochastic Element into Decision Theories," *European Economic Review* 39: 641-648 (1995).

[21] MacKay, David J.C., *Information Theory, Inference and Learning Algorithms,* Cambridge: Cambridge University Press, 2003.

[22] Mackowiak, Bartosz, and Mirko Wiederholt, "Optimal Sticky Prices under Rational Inattention," CEPR discussion paper no. 6243, April 2007.

[23] McFadden, Daniel, "Structural Discrete Probability Models Derived from Theories of Choice," in C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data and Economic Applications,* Cambridge, MA: M.I.T. Press, 1981.

[24] McFadden, Daniel, "Economic Choices," *American Economic Review* 91: 351-378 (2001).

[25] Midrigan, Virgiliu, "Menu Costs, Multi-Product Firms, and Aggregate Fluctuations," unpublished, Federal Reserve Bank of Minneapolis, August 2006.

[26] Miranda, Mario J., and Paul L. Fackler, *Applied Computational Economics and Finance,* Cambridge: M.I.T. Press, 2002.

[27] Moscarini, Giuseppe, "Limited Information Capacity as a Source of Inertia," *Journal of Economic Dynamics and Control* 28: 2003-2035 (2004).

[28] Nakamura, Emi, and Jon Steinsson, "Monetary Non-Neutrality in a Multi-Sector Menu Cost Model," unpublished, Columbia University, October 2007.

[29] Nevin, John A., "Psychophysics and Reinforcement Schedules: An Integration," in M.L. Commons and J.A. Nevin, eds., *Quantitative Studies of Operant Behavior, Vol. 1: Discriminative Properties of Reinforcement Schedules,* Cambridge, MA: Ballinger, 1981.

[30] Paciello, Luigi, "The Response of Prices to Technology and Monetary Policy Shocks under Rational Inattention," unpublished, Northwestern University, October 2007.

[31] Paustian, Matthias, "The Role of Contracting Schemes for the Welfare Costs of Nominal Rigidities over the Business Cycle," unpublished, University of Bonn, April 2005.

[32] Phelps, Edmund S., *Seven Schools of Macroeconomic Thought,* Oxford: Clarendon Press, 1990.

[33] Reis, Ricardo, "Inattentive Producers," *Review of Economic Studies* 73: 793-821 (2006).

[34] Reis, Ricardo, "Inattentive Consumers," *Journal of Monetary Economics* 53: 1761-1800 (2006b).

[35] Rubinstein, Ariel, "Finite Automata Play the Repeated Prisoner's Dilemma," *Journal of Economic Theory* 39: 83-96 (1986).

[36] Shannon, Claude E., "A Mathematical Theory of Communication," *Bell System Technical Journal* 27: 379-423 and 623-656 (1948).

[37] Sims, Christopher A., "Stickiness," *Carnegie-Rochester Conference Series on Public Policy* 49: 317-356 (1998).

[38] Sims, Christopher A., "Implications of Rational Inattention," *Journal of Monetary Economics* 50: 665-690 (2003).

[39] Sims, Christopher A., "Rational Inattention: A Research Agenda," unpublished, Princeton University, March 2006.

[40] Thurstone, Louis L., "A Law of Comparative Judgment," *Psychological Review* 34: 273-286 (1927).

[41] Train, Kenneth E., *Discrete Choice Methods with Simulation,* Cambridge; Cambridge University Press, 2003.

[42] Tutino, Antonella, "The Rigidity of Choice: Lifecycle Savings with Information-Processing Limits," unpublished, Princeton University, December 2007.

[43] Wilson, Andrea, "Sustainable Reputations with Finite Memory," unpublished, Harvard University, December 2006.

[44] Woodford, Michael, *Interest and Prices: Foundations of a Theory of Monetary Policy,* Princeton: Princeton University Press, 2003.

[45] Woodford, Michael, "Information-Constrained State-Dependent Pricing," unpublished, Columbia University, April 2008.

[46] Zbaracki, Mark, Mark Ritson, Daniel Levy, Shantanu Dutta, and Mark Bergen, "Managerial and Customer Costs of Price Adjustment: Direct Evidence from Industrial Markets," *Review of Economics and Statistics* 86: 514-533 (2004).