Memory, Attention and Choice*

Pedro Bordalo, Nicola Gennaioli, Andrei Shleifer

First draft, November 2014. This version, March 2015.

Abstract

We present a theory of consumer choice that combines elements of selective recall and of allocation of attention distorted by salience. The theory helps clarify and organize a variety of evidence dealing with consumer reaction to information, including surprises in quality and prices, unshrouding of hidden attributes such as taxes or maintenance costs, and reminders. Our model explains how consumers under or overreact to information, depending on what draws their attention. It also yields a normative analysis of reaction to reminders which adjusts the "sufficient statistic" methodology.

*Royal Holloway, University of London, Universitá Bocconi, Harvard University. Shleifer thanks the Kauffman Foundation for research support. We are grateful to Paulo Costa, Matt Gentzkow, Sendhil Mullainathan, Josh Schwartzstein, Jesse Shapiro and Jann Spiess.

1 Introduction

Simonsohn and Loewenstein (2006) present a fascinating empirical finding on rent expenditures by movers between American cities. Compare two households, one which moves from San Francisco to Pittsburgh, and another that already lives in Pittsburgh, where rents are 3 times lower. The evidence shows that the former household will spend more on rent than the latter, holding income and several demographic characteristics constant. In turn, a household that moves from a very low rent city to Pittsburgh will spend less on rent than a very similar household living in Pittsburgh already. This effect dissipates over time as households change apartments within cities, but, initially, movers' decision on how nice an apartment to rent is influenced by the rent level in the city they are coming from.

Consider another, apparently very different, finding. A buyer has just agreed to buy a TV for $350 in an appliance store, but has not paid yet. The salesman congratulates the buyer on his good choice, but then reminds him that very occasionally even the best TV's break. In fact, the store offers an insurance policy whereby the TV will be fixed or replaced by a new one if it breaks in the next five years. The cost of this policy is $100. The buyer, imagining the disaster of a broken TV, agrees. In fact, since TV's almost never break in five years, this insurance policy turns out to be vastly overpriced (Sydnor 2010, Huysentruyt and Read 2010). Some evidence indicates that appliance retailers make up to half of their profits from selling such policies (Consumer Reports, 2005).

Although these findings appear unrelated, they have in common that agents' reaction to information is very different from what economic models would predict. In the first example, movers from San Francisco to Pittsburgh remember high rents in San Francisco and thus perceive the rent of high quality apartments in Pittsburgh to be perfectly normal compared to average San Francisco prices. Movers perceive these apartments as good deals, in that their high quality, and not their price, is salient. As a consequence, these movers *under-react* to price differences within Pittsburgh and spend too much on rent. In principle, rents in San Francisco should be irrelevant to their decisions. In the second example, buyers are surprised by the reminder that TV's can break, which they likely have forgotten about since TV's hardly ever break, and *over-react* to this information by purchasing over-priced insurance. Even after being reminded that TV's can break, buyers should be willing to pay only a trivial amount for insurance.

We believe that there is a common psychological mechanism behind under- and over-reaction in the two examples. In both cases, the consumer evaluates his current choice options in light of what is at the top of his mind. His attention is drawn to the aspect along which the contrast is largest. In the mover example, the price of high quality apartments in Pittsburgh is does not stand out relative to average San Francisco prices. On the contrary, the quality of these apartments stands out, leading the consumer to spend too much on rent. In the TV example, the contrast between a broken and a working TV draws attention, and causes the buyer to overpay for insurance (i.e., for a working TV). In both examples, what matters are the aspects of choice that are most distinct from what agents are thinking about or are reminded of. The interaction between memory and attention is necessary to understand the evidence.

In this paper, we present a theory of choice by consumers with selective memory and attention that accounts for this and other evidence. Our theory describes what comes to the consumer's mind when he faces a choice set, what attributes in that choice set surprise him in light of what comes to mind, and how he reacts to the surprises. The theory combines a model of selective recall, following Kahneman and Miller's (1986) Norm Theory, with the theory of attention allocation to unusual or surprising features of choices which we have called Salience Theory (BGS 2012, 2013). Our paper unifies a number of empirical puzzles. It provides new insights into the shrouded attributes model (Gabaix and Laibson 2006), including an account of when firms choose to shroud or unshroud product attributes. Our approach also yields a theory of reaction to reminders, predicts when reminders are more likely to work, and offers an adjustment to the welfare analysis of reminders that has been adopted following Chetty, Looney and Kroft (2009).

The central innovation of the paper is our model of memory, which highlights the associative nature of recall. Formally, we introduce the *availability* function which maps the consumer's past experiences and objective probabilities of attributes and their realisations into the likelihood of recall. We then assume that recall is selective, in that only sufficiently available information about a good comes to mind. This yields two basic departures from a fully rational use of information in memory that are at the core of our paper. First, some information comes to mind that is irrelevant for the choice at hand, as in the example of the mover to Pittsburgh who recalls rental prices from San Francisco. In this case, past prices are available in memory because movers have high exposure to them. Second, some relevant information may fail to be recalled, as in the TV breakdown example. In this case, the breakdown risk is not available in memory because it has low probability and consumers have low exposure to it.

Our main results explore the implications of the interaction between selective memory and attention. Consider first a consumer whose memory is selective but who makes decisions undistorted by the allocation of attention given what he remembers. We call such a consumer Forgetful But Otherwise Rational, or FBOR. In many economic applications, such as Gabaix and Laibson's (2006) theory of shrouded attributes, or Chetty, Looney, and Kroft's (2009) theory of tax salience, consumers are assumed to be FBORs. Although FBORs' decisions can be poor because they forget and do not take into account some relevant aspects of the problem, such as the fact that products have associated maintenance costs or that taxes have to be paid at the register, they correctly incorporate this information into their decisions once reminded. Their reaction to reminders is accurate, and necessarily improves choices.

This formulation appears to miss an important aspect of reality, namely that the consumer's reaction to surprises is often not what an FBOR would do. In our introductory examples, the consumer under-reacts to the rent of high quality apartments in Pittsburgh, and over-reacts to the risk that a TV can break. Consumers also overreact to gasoline price changes, and switch the quality of the gasoline they purchase much more dramatically than would be predicted by income effects (Hastings and Shapiro 2013). Similarly, consumers over-react to overage charges, switching to expensive banking plans that actually raise their average costs (Ater and Landsman 2014). In experimental settings, consumers buy more computer memory when they faced higher prices for memory in previous experimental rounds (Simonson and Tversky 1992). Subjects are willing to pay more for a dinnerware set with 24 intact pieces than for a set with 31 identical pieces plus a few broken ones (Hsee 1998). Subjects overreact to surprising wage increases by providing more effort, compared to subjects receiving identical but expected wages (Gilchrist, Luca and Malhotra 2014). An

FBOR would not show such overreaction.

Nor do reminders that give consumers more information necessarily lead to better decisions. Many of the reminders that bring decision-relevant information to consumers' attention are ignored. Other reminders, such as the fact that a newly purchased TV can break, lead consumers to buy overpriced TV insurance or extended warranties (Huysentruyt and Read 2010). An FBOR always responds to reminders optimally, exhibiting neither underrection nor overreaction. We need a theory that comes to grips with the evidence that consumer reaction to information can be far from optimal.

For this reason, we believe it is important to combine the assumption of selective memory with that of selective attention. The mechanism of salience enables us to examine not just consumer's choice in a static environment, but also his reaction to new information relative to what he recalls. We call such consumers Forgetful And Salient Thinkers, or FASTs, and compare their behavior to that of FBORs. The focus on FASTs enables us to examine how consumers react to surprises. The predictive power of our model stems largely from the interaction between selective recall and selective attention, as exemplified by FASTs.

Table 1 clarifies how our paper compares to some related studies. Rational agent models assume both perfect memory and undistorted processing of information. There is a growing body of research on agents with selective memory but likewise undistorted processing of information, particularly Gabaix and Laibson (2006). This research cannot account for the over-reaction evidence that we have described. In our previous work (Bordalo, Gennaioli, Shleifer 2012, 2013), we considered agents with distorted allocation of attention, but no memory in the sense that the choice context is shaped by their rational expectations. This approach cannot explain how memory influences choice, as in Simonsohn and Loewenstein (2006). By including both selective memory and selective attention, the current model can account for evidence on both over- and underreaction to information and reminders.

In the next four sections we develop this analysis. Section 2 presents the model. Section 3 considers the simplified case in which there are no forgotten attributes, and shows how the model can be used to study the reaction to price surprises. This section works out our account of the Simonsohn-Loewenstein evidence. In section 4, we turn to the case in which decision makers forget some attributes, and study their reaction to information. This section

Attention \ Memory	Perfect	selective
Perfect	Rational Agent	FBOR (Gabaix-Laibson 2006, Chetty at al. 2009)
selective	Salient Thinker (BGS 2013)	FAST (this paper)

Table 1:

shows the fundamental difference in how FBORs and FASTs react to information.

The following sections study the applications. Section 5 investigates the theory, and the growing evidence, on shrouding and unshrouding product attributes, and relates our work to Gabaix and Laibson and the extensive empirical literature that followed. Section 6 studies reminders, beginning with Chetty et al's work on taxes, and continuing with Karlan et al's (2014) work on savings. We also show how to adjust the normative analysis on consumer reaction to information when consumers are FASTs rather than FBORs.

Our paper is related to at least four broad strands of research in psychology and behavioral economics. We follow Kahneman (2011) in formulating the choice problem as beginning with a quick and automatic mental representation. Both recall and allocation of attention are fundamentally system 1 mechanisms. More specifically, in modeling memory we follow Kahneman and Miller's (1986) approach, in which a probe (here a choice set) recruits through imperfect recall "normal" instances of similar options, which like them we call the "evoked set." In turn, attention is allocated to those aspects of choice options that are salient or surprising in light of what comes to mind (BGS 2013).

Second, our paper is related to research on memory, particularly focusing on selective recall of attributes rather than of goods (Mullainathan 2002, Gennaioli and Shleifer 2010, Schwartzstein 2014). There is a growing literature on forgotten goods and actions, such as going to the gym or taking medication, and on how reminders work in that context (DellaVigna and Malmendier 2006, Calzolari and Nardotto 2014, Vervloet et al 2012, Ericson 2014, Taubinsky 2014). We do not pursue this direction here. We model recall through the availability function, which captures the consumer's representation of the normal attributes for each good in the choice set. The availability function takes account of the objective frequency of attributes but also of the consumer's *exposure* to them, which allows incorporation of personal past experience with the good. By incorporating exposure into our model of recall, our work is related to research on anchoring and imprinting of past experience, including Tversky and Kahneman (1974), Gilboa and Schmeidler (1995), and Ariely et al (2003). Our formulation of recall in terms of the availability function enables us to make testable predictions as to what features of the choice options consumers take into account, as well as under- and overreaction to information.

Our formulation of recall and the choice context is related to a large literature on context-dependent choice and reference points in economics, most famously exemplified by Kahneman and Tverksy's (1979) Prospect Theory. Koszegi and Rabin (2006, 2007) substantially advanced this approach by observing that reference points are shaped by expectations, and imposed the assumption of rational expectations to close their model. Others have adopted the view that the reference point is given by recent personal experience, such as recent consumption (DellaVigna et al 2014). We explicitly model recall based in part on forward-looking reality and in part on past experience. Although our formulation is less constraining, it still allows many empirical predictions. For example, it enables us to discuss reference points influenced by rents in a city a person is moving away from, a phenomenon difficult to account for in a forward looking framework. At the same time, in our model recall adjusts to new situations, and the true distribution of prices (even those not experienced personally) affects one's reference point.

Finally, our research is related to the rapidly growing literature on selective attention, and ex post allocation of attention. An early paper exploring how selective attention may distort choices (in ways that might mistakenly be attributed to biases in preferences) is Akerlof (1991). We follow the salience formulation of Bordalo, Gennaioli, and Shleifer; other papers in this area include Koszegi and Szeidl (2013), Gabaix (2014), and Bushong, Rabin and Schwartzstein (2014). The principal advance here is to present a model in which the representation of the problem comes from a combination of memory and attention, and to derive the implications of the interplay between the two.

We should clarify our use of terms, particularly the words "salience" and "attention". Chetty et al. (2009) and Schwartzstein (2014) use "salient" as a synonym for "remembered"', or attended to.[1] Here we distinguish recall from salience, and adopt for the latter the standard definition in Cognitive Psychology, namely that "salience refers to the phenomenon that when one's attention is differentially directed to one portion on the environment rather than to others, the information contained in that portion will receive disproportionate weighing in subsequent judgments" (Taylor and Thompson 1982). We thus refer to things related to memory as remembered, recalled or forgotten, and things that draw a disproportionate share of attention, conditional on what is remembered or observed, as salient.

2 The Model

2.1 Preferences

A consumer is in a choice situation (at a store, restaurant, etc.) in which he must select one of $K > 1$ goods. A generic good k is characterized by quality q_k, upfront price p_k and a "shrouded" attribute f_k, where $(q_k, p_k, f_k) \geq (0, 0, 0)$. We call "shrouded" a non-visible price or quality dimension of the good. A rational consumer observes or recalls the entire choice set $C = \{(q_k, p_k, f_k)\}_{k=1,\ldots,K}$, including the shrouded attribute, and picks the good maximizing his separable utility:

$$\max_k \theta_1 q_k - \theta_2 p_k - \theta_3 f_k. \tag{1}$$

The consumer's income is sufficiently high that his budget constraint does not bind. Different values of parameters $(\theta_1, \theta_2, \theta_3)$ allow the model to capture different cases of interest:

- $\theta_1 = \theta_2 = 1$, $\theta_3 = 0$. There is no shrouded attribute, and goods are represented by

[1] Schwartzstein (2014) considers limited memory in the context of a learning model. A consumer who forgets dimensions he believes have little predictive power for an outcome of interest fails to attend to these dimensions when learning from new data.

quality-price vectors (q_k, p_k), as in BGS (2013).

- $\theta_1 = \theta_2 = 1$, $\theta_3 > 0$. Attribute f_k is a hidden cost. For instance, it could be the cost of replacing a printer's cartridge (Gabaix and Laibson 2006), a sales tax (Chetty et al 2009), or the overage fee in a cellphone contract (Grubb 2011). Parameter θ_3 captures the frequency or probability with which the cost is borne (e.g. the frequency of replacing a cartridge or of incurring overage fees).

- $\theta_1 \geq 0, \theta_2 = 1$, $\theta_3 < 0$. Here f_k is a hidden quality component. For instance, if the good is a home appliance (e.g. a TV), then q_k is the utility of a TV in working order, while f_k denotes its low utility in case of breakdown. In this example, $-\theta_3 = 1 - \theta_1$ reflects the probability that a breakdown occurs.

Attributes q, p and f could in principle be stochastic. For simplicity, we allow only the upfront price p_k to be stochastic, and assume that it is drawn from a (possibly degenerate) probability distribution $\pi_{p_k}(p)$ defined in support X_p. This captures the fact that, at the moment of choice, consumers may face different upfront prices due to cost shocks, sales tactics, or location.[2,3]

Our consumer departs from the rational choice problem (1) due to the combination of selective recall and selective attention. Formally, our consumer follows a three-stage process: first, he observes the visible choice set $C_v = \{(q_k, p_k)\}_{k=1,\ldots,K}$. For example, a consumer buying a TV observes different qualities and prices at a store. C_v acts as a probe for the consumer to recall information that will be relevant to his choice.

In the second stage the consumer i) retrieves from memory the prices that he considers "normal", and ii) may recall the shrouded attribute f_k (e.g. breakdown risk). In recalling the shrouded attribute, the consumer forms his perception of the choice set, which we denote

[2] Distribution $\pi_{p_1}(z)$ should be interpreted as being specific (i.e. conditional) to the choice situation faced by the consumer. For instance, consider the choice situation of buying bottled water: variations in supplier costs or location (e.g. a magazine stand or supermarket) lead to price variability $\pi_{p_1}(z)$. In contrast, buying a bottle of water to accompany a meal at a restaurant is a different need, in the sense that it primes the consumer to condition the price distribution on being at a restaurant, not at the supermarket.

[3] There is a difference between stochasticity in upfront price and other stochastic components of cost. Stochasticity in upfront price refers to different upfront price realisations that are possible ex ante, but whose realization is known at the time of purchase. In contrast, a shrouded cost f may be stochastic in the sense of being a potential future cost, whose realization is uncertain at the time of purchase.

$\widetilde{C}$. If the consumer does recall the hidden attribute, he includes it in his perceived choice set, which becomes the true choice set, namely $\widetilde{C} = C = \{(q_k, p_k, f_k)\}_{k=1,\dots,K}$. If the consumer does not recall the hidden attribute, his perceived choice set is the visible choice set, $\widetilde{C} = C_v = \{(q_k, p_k)\}_{k=1,\dots,K}$, and the shrouded attribute is neglected in choice.

In the third and final stage, the consumer evaluates the available goods, but his attention is selectively drawn to the attributes that are salient or stand out in the choice set, as well as relative to normal prices. As a consequence, selective recall sets the stage for two types of framing effects that shape the allocation of attention: i) those associated with reminding consumers of a previously neglected shrouded attribute f_k, and ii) those associated with observing a realized price p_k that is either very different or very close to the consumer's normal price p_k^n.[4]

2.2 Imperfect Recall and the Evoked Set

At the recall stage, the consumer represents the choice set as C if he recalls the shrouded attribute f, and as C_v otherwise.[5] The consumer also recalls some of the price realizations for each good k in the choice set.

Formally, in each instance recall acts upon a set X of objects: upon the shrouded attributes, in which case $X = \{f\}$, and upon the price realisations of each good, in which case $X = X_p$ is the common price support of goods in C. Each element $x \in X$ is associated with an "availability" score $a(\pi_x, e_x) \geq 0$ so that ease of recall increases with availability. Here, π_x is the objective frequency of element x in X: when recall acts on the space of shrouded attributes $X = \{f\}$, π_x equals the frequency with which attribute f takes non-zero values, namely $\pi_f = |\theta_3|$. When recall acts on prices, $X = X_p$, where X_p is the price support, the relevant frequency for each possible price $p \in X_p$ for good k is pinned down by the price distribution π_{p_k}.

The second argument in the availability function is the consumer's relative exposure e_x

[4] Stochasticity in q would allow for the second type of surprises to affect the salience of quality. The analysis of quality surprises is entirely analogous to that of price surprises. We thus suppress stochasticity in quality in order to reduce the number of cases that must be analyzed.

[5] We assume that recall of attribute f applies equally to all goods in the choice set. While the ease of recall of attributes may in principle vary across goods (see below), it is realistic to assume that recalling an attribute for one good prompts consumers to consider that attribute for all goods in C.

to $x \in X$. Exposure captures the relative intensity of personal past observation, mental rehearsals, or reminders about x as a share of all instances in which the consumer thought about the good.

In line with existing work in memory research, we impose the following properties:

Definition 1 *The availability function $a(\cdot,\cdot)$ is weakly increasing in both of its arguments, $a_\pi \geq 0$ and $\partial_e a \geq 0$, and satisfies $a(0,0) = 0$, $a(1,1) = 1$. If $\partial_e a \equiv 0$ then $a(\pi, e) = \pi$.*

An example of an availability function satisfying Definition 1 is $a(\pi, e) = \alpha\pi + (1-\alpha)e$. Here, $\alpha \in [0,1]$ measures the strength of exposure in determining availability, and thus recall.

Both frequency and exposure are central ingredients in recall. The assumption that the availability of an element x increases in the consumer's exposure to x captures the idea that mental rehearsal of information facilitates its recall, consistent with a long tradition in Cognitive Psychology (Atkinson and Shiffrin 1969, Jacoby 2001) and with recent work in economics (Mullainathan 2002, Taubinsky 2014). It also implies that repeated exposure to an element x facilitates its recall, independently of its actual frequency or relevance, an idea that plays a role in advertising. Exogenous interventions such as reminders about shrouded attributes can also be conceptualized as increasing the consumer's exposure to them. In the field, exposure e_x can be measured as past personal experience of some realisations of x, as in Malmendier and Nagel's (2009) work on stock market returns, or by differential exposure to advertising. Critically, the role of past exposure implies that recall, and thus beliefs (see below), are at least in part backward looking.

The assumption that availability increases in objective probability π_x captures the idea that recall is also shaped by forward looking elements, in line with the use of information by a rational consumer: knowledge that a price realization or a cost contingency is likely to occur facilitates its recall. If the consumer goes to an upscale supermarket, he expects to face higher prices, even if he has limited past exposure to them.

We model recall as follows:

Definition 2 *There is an availability threshold $a^* \in (0,1]$ such that the consumer recalls the subset $X(a^*) \equiv \{x \in X : a(\pi_x, e_x) \geq a^*\}$. To each $x \in X(a^*)$ he associates the relative availability weight $a(\pi_x, e_x)/\sum_{X(a^*)} a(\pi_y, e_y)$.*

This recall process nests several models used in previous work. The benchmark case of unlimited memory and rational expectations arises when recall satisfies two conditions: i) the availability threshold is not binding, $a^* = 0$, so that all elements in X are recalled, and ii) availability it is not affected by exposure, $\partial_e a \equiv 0$, so that recalled elements are weighted according to their objective frequency (since, from Definition 2 we then have $a(\pi_x, e_x) = \pi_x$ for all $x \in X$).

If exposure does not matter, $\partial_e a \equiv 0$, but memory is selective, $a^* > 0$, consumers may neglect some elements x, but have in mind the correct (conditional) distribution for the recalled elements in $X(a^*)$. Several recent papers use this likelihood-based model of recall (Gennaioli, Shleifer and Vishny 2012).[6] In this case, the less selective is recall (the lower is a^*), the closer the consumer is to the benchmark of rational expectations. At the opposite extreme, where $a^* > a(\pi_x, e_x)$ for all $x \in X$, so that nothing is recalled, $X(a^*) = \emptyset$, the consumer sees the visible attributes in C_v and fails to remember the shrouded cost or alternative prices, in line with the WYSIATI principle (Kahneman 2003).

Finally, if exposure does matter, $\partial_e a > 0$, departures from rational expectations also occur because individuals are swayed by personal experience. Attributes such as taxes which we experience indirectly may not be recalled, even if they are very likely. In the extreme case where $a(\pi_x, e_x) \equiv a(e_x)$, availability depends exclusively on past exposure, as in Mullainathan (2002) and Taubinsky (2014). In this case, a less selective recall process (lower a^*) facilitates the recall of potentially irrelevant information, leading to larger departures from the rational expectations benchmark.

Both exposure and frequency are required to make sense of available evidence on recall. Consider the case of shrouded attributes: attribute f is recalled if $a(\pi_f, e_f) > a^*$, in which case $X(a^*) = \{f\}$, and forgotten otherwise, in which case $X(a^*) = \emptyset$. The presence of a threshold $a^* > 0$ implies that information with low exposure or low frequency is forgotten. For instance, neglect of rare quality or price contingencies such as breakdown risk or overage fees can be due to their low probability π_f. A pure likelihood based recall process, however, cannot explain why consumers sometimes forget attributes that occur with certainty, such

[6] Note a central difference from models of rational inattention, in which neglected information depends on the level of expected payoffs, and not on the likelihood alone.

as sales taxes or maintenance costs (e.g. replacing a printer's cartridge). Here, $\pi_f = 1$, but these attributes can still be forgotten because exposure is so low that $a(1, e_f) < a^*$. Likewise, attributes that receive higher exposure, for example through advertising or repeated rehearsal by the consumer, become more available.

The combination of frequency and exposure is also key in recall of price realisations. When they move to a less expensive city, consumers may recall some low prices, because they are more frequent there. At the same time, they also recall the higher prices they were exposed to in the past, even though these prices are not relevant to their current choice.

In light of the choice set $\widetilde{C}$ perceived by the consumer, recall of attributes and prices give rise to the consumer's "evoked set" as follows:

Definition 3 *Upon seeing the visible choice set C_v, the consumer recalls an evoked set consisting of goods at normal prices:*

$$\widetilde{C}(n) = \begin{cases} \{(q_k, p_k^n, f_k)\}_{k=1,\ldots,K} & \text{if} \quad a(\pi_f, e_f) \geq a^* \\ \{(q_k, p_k^n)\}_{k=1,\ldots,K} & \text{otherwise} \end{cases},$$

where normal prices are equal to:

$$p_k^n = \begin{cases} \sum\limits_{X_{p_k}(a^*)} p \cdot \frac{a(\pi_p, e_p)}{\sum\limits_{X_{p_k}(a^*)} a(\pi_p, e_p)} & if \quad X_{p_k}(a^*) \neq \emptyset \\ p_k & if \quad X_{p_k}(a^*) = \emptyset \end{cases}.$$

Following BGS(2013), we call $\widetilde{C} \cup \widetilde{C}(n)$ the *choice context*. Selective recall shapes the representation of goods in the choice context in two ways: it may prune some utility carriers (if the shrouded attribute is not sufficiently available), and it may also prune which price realizations are considered when forming a norm for the purchase price.[7]

The normal price p_k^n combines all recalled prices, weighting each of them by its availability $a(\pi_{p_k}, e_{p_k})$ (normalised by the total availability of recalled prices). When $\partial_e a = 0$, recalled prices are weighted by their true probabilities, and so the normal price is the conditional ex-

[7]It is straightforward to extend the definition of normal attribute level to the case of hidden attributes.

pectation $\mathbb{E}_{\pi_{p_k}}\left[p_k \,|p_k \in X_{p_k}(a^*)\right]$. The difference between the price that is considered normal and the expected price in $X_{p_k}(a^*)$ reflects the role of exposure in determining p_k^n. When no specific price is recalled ($X_{p_k}(a^*) = \emptyset$), the current price is viewed as normal. In the absence of retrieved memories, the choice context is fully shaped by the choice set.

2.3 Choice Context and Salience

Following BGS (2013), we formalize selective attention by postulating that the consumer does not maximize his true utility in (1). Instead, he maximizes a perceived utility obtained by inflating for each good $k \in \widetilde{C}$ the weight attached to the attribute (quality, price, or shrouded) whose level is salient, in the sense of being furthest from the reference attribute levels in $\widetilde{C} \cup \widetilde{C}(n)$.

Our modelling here is identical to our prior work (BGS 2013). We define the reference level of an attribute as the average value of that attribute in the choice context $\widetilde{C} \cup \widetilde{C}(n)$. The reference quality, price and shrouded attribute are thus respectively given by $\overline{q} = \sum_{k \in C} \frac{q_k}{K}$, $\overline{p} = \sum_{k \in C} \frac{p_k^n + p_k}{2K}$ and $\overline{f} = \sum_{k \in C} \frac{f_k}{K}$.

Salience is measured by a function $\sigma(.,.)$, such that the salience of price p_k for good k is $\sigma\left(p_k, \overline{p}\right)$, that of quality q_k is $\sigma\left(q_k, \overline{q}\right)$ and that of the shrouded attribute f_k is $\sigma\left(f_k, \overline{f}\right)$. The salience of an attribute increases in its *proportional distance* from its reference level. Formally, $\sigma\left(\cdot, \cdot\right)$ satisfies ordering and homogeneity of degree zero. According to ordering, if an interval $[x, y]$ is contained in an interval $[x', y']$, then $\sigma\left(x, y\right) < \sigma\left(x', y'\right)$. According to homogeneity of degree zero, $\sigma\left(\alpha x, \alpha y\right) = \sigma\left(x, y\right)$ for any $\alpha > 0$, with $\sigma(0, 0) = 0$.[8]

Let r_{v_k} be the salience ranking of attribute v for good k in the perceived choice set

[8]Ordering and homogeneity of degree zero of the salience function imply that the salience of a good's quality is an increasing function of the percentage difference between the good's quality and the average quality in the choice context, and similarly for price (see BGS 2013). In particular, consumers have diminishing sensitivity to attribute differences: increasing the prices of both goods by a uniform amount ϵ makes prices weakly less salient, $\sigma(p_k + \epsilon, \overline{p} + \epsilon) \leq \sigma(p_k, \overline{p})$ for $k = 1, 2$, and strictly so when $p_k \neq \overline{p}$. This property is consistent with Weber's law of sensory perception.

$\widetilde{C} \cup \widetilde{C}(n)$. Then, the salient thinker's perceived utility from good k is:

$$u^S(q_k, p_k, f_k | \widetilde{C} \cup \widetilde{C}(n)) = \begin{cases} b_{vk}\left[\theta_1 \delta^{r_{q_k}} q_k - \theta_2 \delta^{r_{p_k}} p_k\right] & \text{if} \quad \widetilde{C} = C_v \\ b_k\left[\theta_1 \delta^{r_{q_k}} q_k - \theta_2 \delta^{r_{p_k}} p_k - \theta_3 \delta^{r_{f_k}} f_k\right] & \text{if} \quad \widetilde{C} = C \end{cases} \tag{2}$$

Less salient dimensions are discounted by $\delta \in [0, 1]$ *relative* to more salient dimensions. (The positive constants b_{vk} and b_k are such that the sum of the absolute values of attribute weights stays constant)[9]. In the case of goods (q_k, p_k) with one quality and one price dimensions, we have

$$u^S(q_k, p_k | C \cup C(n)) = \begin{cases} \frac{2}{1+\delta}\left[q_k - \delta p_k\right] & \text{if} \quad \text{quality } q_k \text{ is salient} \\ \frac{2}{1+\delta}\left[\delta q_k - p_k\right] & \text{if} \quad \text{price } p_k \text{ is salient} \\ q_k - p_k & \text{if} \quad \text{equal salience} \end{cases} . \tag{3}$$

The consumer picks in $\widetilde{C}$ his preferred good $k^* = \text{argmax}_{k \in \widetilde{C}} u^S(q_k, p_k, f_k | \widetilde{C} \cup \widetilde{C}(n))$. In this analysis, parameter δ plays a key role. When $\delta = 1$, the consumer does not pay disproportionate attention to salient dimensions. We call such a consumer an FBOR: Forgetful But Otherwise Rational. An FBOR may be surprised after seeing an unusual price but his behavior is not swayed by the surprise; he continues to choose by maximizing his undistorted utility function (1). When an FBOR observes all relevant product attributes, he behaves like a fully rational consumer.

When $\delta < 1$, our consumer is FAST: Forgetful And Salient Thinker. The allusion to Kahneman (2011)'s system 1 thinker is not accidental. FAST reacts to available information by paying disproportionate attention to salient data. Surprises from unusual price realizations may relocate his focus of attention to quality or price. The effects of salience becomes stronger the smaller is δ. As we show in the following sections, the interaction between selective memory and selective attention of FASTs can account for field and experimental evidence of over- and under-reaction to information that cannot be accounted for with only the selective recall of FBORs.

[9] Formally, $b_{vk} = (\theta_1 + \theta_2)/(\theta_1 \delta^{r_{q_k}} + \theta_2 \delta^{r_{p_k}})$ and $b_k = (\theta_1 + \theta_2 + |\theta_3|)/(\theta_1 \delta^{r_{q_k}} + \theta_2 \delta^{r_{p_k}} + |\theta_3| \delta^{r_{f_k}})$.

2.4 Discussion

In our model, consumers pay disproportionate attention to salient aspects of their choice context. Relative to our previous work, the new ingredient is that the choice context is partially determined by consumer recall. Specifically, i) consumers may fail to consider some attributes, and ii) among the attributes that are top of mind, salience is measured relative to a reference level that is shaped by the choice set as well as by recalled normal prices. The role of recall in shaping context is modelled through the Evoked Set $C(n)$. This captures a longstanding idea in psychology: objects or events serve as cues that trigger consumer recall of their "normal" versions, in turn influencing judgments and choices (Kahneman and Miller 1986).

We introduce this idea into an economic model of decision making. In so doing, we offer a new memory-based approach to modeling reference points. Because reference prices incorporate normal prices, they are shaped by availability and hence by both frequency and exposure. As a consequence, reference points in our model combine forward looking elements (frequent prices), with backward looking elements (past prices with high exposure), and thus include personal experience but may in principle also include observations of others' experiences. In a steady state where exposure is equal to frequency, $e_x = \pi_x$, normal (and thus reference) prices are identical or close to rationally expected prices, in the spirit of Koszegi and Rabin (2006).[10] When however there is an abrupt change in the environment (i.e. a person moves from one city to another), $e_x \neq \pi_x$, and reference prices deviate from rational expeectations, incorporating backward looking aspects. This approach can provide a unified and potentially testable model of reference prices, based on the psychology of availability.

Reference attributes shape valuation by determining which attribute is salient for each good. To formalize this mechanism in a tractable and disciplined way, we have made several modelling choices that we now discuss.

The recall process encoded in the availability function $a(\pi, e)$ incorporates four main

[10] When $e_x = \pi_x$, if memory is unlimited ($a^* > 0$), normal prices are equal to rationally expected prices provided the availability function is homogeneous of degree one, namely $a(k\pi_x, ke_x) = ka(\pi_x, e_x)$ for any $k > 0$. We discuss elsewhere (BGS 2013), our reference point is pinned down by the choice environment, and not by expected consumption as in Koszegi and Rabin (2006).

simplifying assumptions. First, we do not allow for memory inhibition, whereby recall of an object crowds out the recall of other objects. Inhibition is an important feature of recall, but it is not important for our results, so we abstract from it.[11] Second, selective recall truncates away the least available memories. We could allow for a smoother discounting of low availability types, but it seems realistic to assume that consumers fail to recall everything. In fact, this feature allows for forgotten attributes. Third, the model does not capture recall based on the representativeness heuristic, which is important in reality (Kahneman and Tversky 1973). Based on our prior work, we view representativeness as a mechanism suited to describe contrasting probabilistic hypotheses (GS 2010) or situations/social groups (BGS 2014). Finally, our model is static, and so disregards the effect of recency on availability.[12]

A second set of assumptions concerns the formation of the evoked set and the reference prices. The model assumes that recalled prices are weighted by using their relative availability, in line with Tversky and Kahneman (1973). Availability based weighting allows both frequency and exposure to affect normal prices. When moving to a less expensive city, the consumer adjusts his normal price downward, but this adjustment might be insufficient. As a result of this averaging process, the normal price need not be among the recalled prices. Our results hold qualitatively for more general assumptions on the averaging of recalled prices.

With respect to the reference point, we extend Bodner and Prelec (1994) in assuming that the reference level of an attribute is its average level, or " centroid", in the broader context $\widetilde{C} \cup \widetilde{C}(n)$. That is, the reference point aggregates information from the choice set as well as from memory. This is consistent with both adaptation theories of the reference point (in which past or expected prices set the reference) and with context effects of the choice

[11] Because in our model exposure is measured as a frequency, it features some memory inhibition: increasing exposure to one price decreases exposure, and availability, of other prices. The model could be extended to also feature a more conventional form of memory inhibition by adding the constraint that the total availability of recalled elements cannot exceed a threshold $a^{**} > a^*$, namely the set of recalled elements consists of the set $X(a^*, a^{**})$ of the most available elements $x \in X$ such that:

$$a(\pi_x, e_x) \geq a^*, \ \sum a(\pi_x, e_x) \leq a^{**}.$$

The second constraint captures inhibition, for it implies that recall of a highly available items crowds out recall of less available items.

[12] Recency effects have been shown to affect learning and shape beliefs (Hogarth and Einhorn 1992, Baucells, Weber and Welfens 2011), and are likely of central importance for advertising and the effect of reminders, see Ericsson (2014), Taubinsky (2014). While we do not model it here, we note that the current framework could easily be extended to incorporate recency effects in the availability function.

set itself (as typified by the decoy effect). This approach does not explicitly link reference points to expectations, even though the psychological notions of normal price and expected price are closely related. We leave this question for future work.

Finally, our model describes choice settings in which consumers actively consider a (typically small) number of possible alternatives that are relevant to their needs, and which can in principle be known to the modeler. This notion of choice set is related to that of *consideration sets* in the marketing literature (Shocker et al. 1991). A significant literature in marketing and economics has explored the formation of consideration sets both empirically (Nedungadi 1990) and theoretically, due to consumers' cognitive limitations (Hauser and Wernerfelt 1990, Roberts and Lattin 1991), or to manipulation by firms (Eliaz and Spiegler 2011). We do not endogenize choice sets in this paper, although this is an important direction for future research.

3 Availability, Normal Prices and Choice

We now illustrate our model in the simplest setting in which there are only two product attributes, quality and price, and $\theta_1 = \theta_2 = 1$ and $\theta_3 = 0$ in Equation (1). Both quality and price are visible attributes. Thus, selective memory only affects price expectations, which in turn shape the consumer's attention.

We frame the problem in the context of Simonsohn and Loewenstein (2006)'s findings, namely that, controlling for observables including income, US movers coming from more expensive cities rent more expensive apartments than those coming from cheaper cities and vice versa. Relatedly, Simonsohn (2006) documents that commuting times of movers to cities tend to be biased towards those of the city of origin.

These papers provide field support for what Simonson and Tversky (1992) called "background context effect", whereby choice sets encountered at the beginning of an experiment affect subjects' subsequent choices.[13] This phenomenon is related to the widely documented

[13] In their experiment, subjects choose between computers differentiated in terms of memory and price. The marginal price of computer memory is given implicitly by the variation in choice set along the price and memory dimensions, and can be high, intermediate or low. Subjects who were first exposed to a choice set in which the implied price of memory is high subsequently choose to purchase more memory in the intermediate-price setting than do subjects who were first exposed to a low-price setting.

contrast effects *within* the choice set, but is distinct because contrast is shaped by memories. In our model, the interaction between memory and attention produces background context effects: consumers think about features of earlier choice sets (i.e., prices), which in turn shape what they attend to in the current situation, and affect their choices.

3.1 Normal Prices and Choice

A consumer is looking to rent an apartment in *Pit* (for Pittsburgh). There are apartments of two qualities, $q_1 > q_2$, at prices $p_1 > p_2$, so $C = \{(q_1, p_1), (q_2, p_2)\}$. Assume that a rational consumer is willing to rent and is indifferent between the two apartments, namely $q_1 - p_1 = q_2 - p_2 > 0$ (which also implies that $q_1/p_1 < q_2/p_2$). This is a convenient simplification, but the analysis is more general (see Appendix 1).

Faced with the choice set C, the consumer recalls an evoked set $C(n) = \{(q_1, p_1^n), (q_2, p_2^n)\}$, which includes the apartments at the prices he considers normal. Suppose that normal prices display a parallel shift with respect to actual prices, namely they satisfy $p_k^n = p_k + \Delta^n$, for $k = 1, 2$, where $\Delta^n > -p_2$. When $\Delta^n > 0$ the consumer views the current average rent as below normal. When $\Delta^n < 0$ the consumer views the current average rent as above normal. Normal prices are taken as given for now; we endogeneize them in section 3.2.

The consumer's view on normal prices renders the evoked set, and thus the choice context, consumer specific. The reference quality is $\overline{q} = (q_1 + q_2)/2$ and the reference price is $\overline{p} = (p_1 + p_2 + \Delta^n)/2$. For this consumer, then, the quality of apartment $k = 1, 2$ is salient relative to its rental price if and only if:

$$\sigma\left(q_k, \frac{q_1 + q_2}{2}\right) > \sigma\left(p_k, \frac{p_1 + p_2 + \Delta^n}{2}\right). \tag{4}$$

Whether the consumer sees the quality or the price to be salient for a given apartment depends on how the apartment's rent compares to the reference rent $\overline{p}$ in the choice context $C \cup C(n)$. In turn, this depends on the average normal rent perceived by the consumer.

According to our model, this result requires that the consumer puts some weight on the prices of the first stage when constructing the "normal" price for the second stage. This is consistent with our specification as long as the first price lies in the support of the price distribution for the second stage. More generally, the result obtains in a straightforward way if the normal price is the availability-weighted (instead of the frequency-weighted) average of prices that come to mind.

Consider first the case where the high quality apartment in Pittsburgh is weakly more expensive than average, namely $p_1 \geq \overline{p}$. This is satisfied provided normal prices are not too high relative to the price dispersion within Pittsburgh, namely $\Delta^n < p_1 - p_2$. It then follows from (4) (and homogeneity of degree zero of the salience function) that the quality of this apartment is salient if and only if its quality-price ratio is higher than average, that is

$$\sigma\left(q_1, \frac{q_1 + q_2}{2}\right) > \sigma\left(p_1, \frac{p_1 + p_2 + \Delta^n}{2}\right) \quad \Leftrightarrow \quad \frac{q_1}{p_1} > \frac{\overline{q}}{\overline{p}} \tag{5}$$

There are thus two cases. Equation (5) holds whenever $\Delta^n > \Delta^{n*} = p_1\left(\frac{q_2}{q_1} - \frac{p_2}{p1}\right)$, which by assumption is positive. Intuitively, if normal rents are sufficiently higher than rents in *Pit*, then the reference rent $\overline{p}$ is close to the rent p_1. Apartment 1 is then perceived as a "good deal", in the sense that it provides higher than average quality for an average price. The consumer's attention is therefore drawn to the apartment's higher quality, so $u^S(q_1, p_1) = \frac{2}{1+\delta}(q_1 - \delta p_1)$. This implies that apartment 1 is overvalued relative to apartment 2, and in particular the consumer undervalues the rent differential across apartments, namely $\frac{2}{1+\delta}(\delta p_1 - p_2) < p_1 - p_2$. In other words, the consumer *underreacts* to the rent difference.

If instead $\Delta^n < \Delta^{n*}$, then the reverse of equation (5) holds: in this case, normal rents and thus the reference rent are low, so the high rent p_1 is salient. Apartment 1 is then perceived as a "bad deal", and is valued at $u^S(q_1, p_1) = \frac{2}{1+\delta}(\delta q_1 - p_1)$. As a consequence, the consumer overvalues apartment 2 relative to apartment 1, and in particular overvalues the rent differential, namely $\frac{2}{1+\delta}(p_1 - p_2) > p_1 - p_2$. In other words, he *overreacts* to the rent difference.

Finally, consider the case where normal rents are so high that the reference price $\overline{p}$ is higher than p_1. In this case, the expensive apartment's quality is salient if and only if $q_1 \cdot p_1 > \overline{q} \cdot \overline{p}$, which holds whenever $\Delta^n < \Delta^{n**} = p_1 \cdot \frac{3\frac{q_1}{q_2} - 1}{\frac{q_1}{q_2} + 1} - p_2$. Intuitively, 1's higher quality is salient provided $\overline{p}$ is "close" to p_1; but if the recalled prices are sufficiently high relative to the price dispersion in Pittsburgh, then the difference between choice set prices and evoked set prices becomes very salient. The consumer then focuses on the low prices of both apartments in Pittsburgh, and as a consequence overvalues the cheaper apartment.

Note the key role of memory: recall of different normal price realizations can direct attention either to the quality of the high end apartment or to its price. In the first case, the consumer under-reacts to price information, while in the second he over-reacts, causing valuation to change drastically.

The choice behavior of the consumer is characterized as follows:

Lemma 1 *Let* $\Delta^{n*} = p_1 \left(\frac{q_2}{q_1} - \frac{p_2}{p1} \right) > 0$ *and* $\Delta^{n**} = p_1 \cdot \frac{3\frac{q_1}{q_2} - 1}{\frac{q_1}{q_2} + 1} - p_2 > p_1 - p_2$*. Then:*

i) for $\Delta^n < \Delta^{n*}$ *price is salient for apartment 1, and the consumer chooses apartment 2.*

ii) for $\Delta^n \in (\Delta^{n*}, \Delta^{n**})$ *quality is salient for 1, and the consumer chooses it.*

iii) for $\Delta^n > \Delta^{n**}$ *price is salient for 1, and the consumer chooses 2.*

The Lemma shows that consumers with the same preferences but with different price norms react differently to the same choice set. Even though apartments 1 and 2 provide exactly the same rational utility, a high price norm consumer tends to spend more on housing than a low price norm consumer. This difference stems from the consumers' different reactions to price realisations, which are perceived in light of the normal prices.

We now use our model of recall to endogenize the reference prices p_k^n. By doing so, we illustrate under which conditions our model yields the background contrast effects, and in particular the Simonsohn and Loewenstein evidence.

3.2 Recall and Normal Prices

When faced with actual rental prices, consumers spontaneously retrieve from memory available price realizations. There are two types of consumers, identified by their histories: those coming from Pittsburgh and those coming from San Francisco. Denote the type of consumer by $j = Pit, SF$. For consumer j, the availability of price p for apartment $k = 1, 2$ in Pittsburgh is $a(\pi_{p_k}^{Pit}(p), e_{p_k}^j(p))$ where $\pi_{p_k}^{Pit}(p)$ equals 1 if $p = p_k$ and 0 otherwise. The frequency component of recall is common to all consumers searching in the same city, and concentrated on the actual price p_k of apartment k in Pit. The exposure component of recall, in contrast, may vary across consumers, as it can depend on their past price experiences (see Section 2.2).

To formalise the idea that consumers' histories affect their notion of normal prices, we assume that price exposure equals the price frequency observed in the past (the results go through under the more relaxed assumption that exposure merely increases in past frequency). Consider first the consumer $j = Pit$, accustomed to Pittsburgh. In his case, exposure and frequency coincide. As a consequence, for each apartment k availability is maximal at the true price and minimal otherwise, namely p_k has availability $a(1, 1)$ while $p \neq p_k$ has availability $a(0, 0)$.

Matters are different for a mover to Pittsburgh from San Francisco, namely $j = SF$. In his case, immediately after moving, exposure differs from frequency. Indeed, suppose that in San Francisco the price of an apartment k is $p_k + \Delta$. Then, when thinking about the same quality in Pittsburgh, the availability of this higher San Francisco price is $a(0, 1)$. This is due to the fact that $p_k + \Delta$ has zero frequency in Pit but full exposure in the consumer's mind. At the same time, for this consumer the availability of the current Pittsburgh price p_k is equal to $a(1, 0)$: the consumer knows that it is the current price, but has zero personal exposure to it. All other prices have minimal availability $a(0, 0)$. The mover to Pittsburgh might then recall San Francisco prices, particular if exposure is more important than frequency in shaping recall.

We have:

Lemma 2 *Suppose that both consumers SF and Pit search in city Pit. Then, for all a^* we have $p_k^{n,Pit} = p_k$. If $a^* \leq a(0, 1)$ then $p_k^{n,SF} > p_k$, while if $a^* > a(0, 1)$ then $p_k^{n,SF} = p_k$.*

If memory is very limited, $a^* > a(0, 1)$, the mover immediately forgets his prior experiences, and immediately perceives Pittsburgh rents as normal. In this case, both SF and Pit perceive the same normal rents and behave in the same way. By Lemma 1, they both rent the low quality apartment 2.

When instead memory is not so limited, $a^* < a(0, 1)$, then prior experience shapes the rents consumers view as normal. In this case, the more (irrelevant) information is recalled, the more biased are normal prices relative to a rational expectations benchmark. Because consumer SF recalls the higher rents in San Francisco, he perceives a higher normal rent than the native from Pittsburgh. According to Lemma 1, for a broad range of high recalled prices, consumer SF perceives the expensive apartment 1 as a good deal and chooses it.[14]

[14] Lemma 1 states that if rent differences across cities are sufficiently large relative to the dispersion of

Differences in prior experience, though normatively irrelevant, translate into differences in behavior. This logic helps account for the Simonsohn and Loewenstein (2006) evidence: when consumer SF moves to the cheaper city Pit, he views local high quality apartments as good deals, and ends up renting one of them. Instead, the local consumer Pit views high quality apartments as too expensive in light of his lower price norms. Similarly, when a consumer moves to Pit from Tuscon, where rents are even cheaper, he views local apartments as too expensive. Given a broader choice set in Pittsburgh, this mover would rent a lower quality apartment than a local.

History matters for behavior because, by shaping consumers' idea of normal prices, past exposure shapes attention allocation, and thus valuations and choices. Obtaining the background context effect in our model requires both exposure-driven recall (as highlighted in Lemma 2) *and* salience distortions (as in Lemma 1). Without exposure-driven recall, $\partial_e a = 0$, past experience would not shape normal prices. We would be in a rational expectations world, in which SF and Pit consumers behave in the same (distorted) way. Without salience, $\delta = 1$, normal prices would not distort attention allocation. As a consequence, price norms would not affect behavior, so that SF and Pit consumers again behave identically.[15]

Critically, our model makes predictions on how the behavior of movers SF changes over time. As rents in Pit become more available to movers through repeated exposure, and rents in SF become less available through diminished exposure, the norm of movers converges towards the norm of stayers. This fosters adaptation, rendering the choice behavior of

rents within each city, namely if $p_1^{n,SF} - p_1 > p_1 \frac{3q_1/q_2 - 1}{q_1/q_2 + 1} - p_2$, the mover focuses on prices and reverts to the cheap apartment. This condition, however, becomes harder to fulfil if more than two apartments are considered, in such a way that the dispersion of realized prices increases while $p_1^{n,SF} - p_1$, and thus the reference price, stay constant. Intuitively, while prices for similar apartments may vary dramatically across different cities, there typically exists a sufficient spread of qualities, and thus of prices, within each city that the full price distributions have significant overlaps. In this case, it can be shown that chosen quality is monotonically increasing with recalled prices.

[15]Evidence suggests that past personal experience carries disproportional weight in current decisions. Hastings and Shapiro (2013) show that when the cost of gasoline increases, consumers substitute away from the high octane (expensive) gas to an extent that is hard to reconcile with income effects. This pattern is consistent with a salient thinker whose normal price is the lower, historical price of gas. Past prices (such as the purchase price) has also been shown to influence consumers' behavior, in the housing market (Genovese and Mayer, 2001), the stock market (Odean 1998) and the labor market (DellaVigna et al. 2014). Malmendier and Nagel (2009) show that returns experienced by an investor in the stock market affect the extent of risk taking in future investments. While our analysis does not explicitly incorporate recency effects, these findings are consistent with our model under the assumption that recent experiences are more available in memory.

SF indistinguishable from that of Pit. This prediction is confirmed in Simonsohn and Loewenstein (2006), who find that the second house purchase of movers is indistinguishable (on average) from the purchases of natives.[16]

So far, we have considered the case where exposure differs from frequency due to a sudden change in the environment. However, exposure and frequency might differ for other reasons as well (the special case where exposure coincides with frequency is examined in more detail in Appendix B). For example, exposure may be driven by factors other than actual price observations, such as advertising. An advertising campaign about a firm's sales event may increase exposure to sale prices, over and above their actual frequency, leading consumers to represent this brand as being normally cheaper than the competitors. This logic suggests an intuitive role for advertising, namely that of *breaking through the availability threshold*, and populating the consumer's evoked set with normal attributes that draw the consumer's attention to the advantages of the advertised product. Another interesting case is when exposure changes dramatically in response to a change in frequency. For example, the spread of an infectious disease, or a surge in crime, are likely to be heavily reported by the media, boosting exposure to such contingencies over and above the actual risk of their occurrence (even if the latter has in fact increased a bit). In these cases, our model suggests that changes in behavior will be excessive.

4 Forgotten Attributes and Reaction to Information

As we saw in Section 2, consumers may fail to retrieve certain attributes from memory when they are shrouded (not visible in the choice setting) and insufficiently available. We now show that the interaction between memory and attention can have far-reaching consequences for the way in which FASTs react to information about neglected attributes. Reminding the

[16]Simonsohn and Loewenstein's finding that movers differ from natives on impact might also be consistent with loss aversion. However, this requires that the reference price is anchored at prices in the city of origin, and thus not determined by expectations, counter to Koszegi and Rabin (2006). If loss aversion were instead defined relative to past consumption, the evidence would suggest that loss aversion operates more strongly on prices than on qualities. Importantly, loss aversion relative to past housing experience cannot explain the adaptation pattern by movers, as it would imply that movers are reluctant to relocate from their initial choice of house in the new city. For a detailed discussion of the differences between loss aversion and the salience mechanism, see BGS (2012, 2013).

consumer of a neglected dimension has two effects. First, by increasing the consumer's information, it changes the perceived choice set from C_v to C (and similarly for the evoked set). Second, it can affect the agent's allocation of attention to goods' attributes. The second effect can generate under or over-reaction to information, relative to an FBOR.

Consider the following simple example. A consumer chooses between two laptops, characterized by their qualities q_k, upfront prices p_k and maintenance costs f_k. The true utility of laptop k is

$$q_k - p_k - \mu f_k. \tag{6}$$

where $\mu > 0$ measures the "rate of maintenance," e.g. how often components must be replaced.[17] Laptop 1 has weakly higher quality $q_1 \geq q_2$ and costs $p_1 \geq p_2$, $f_1 \geq f_2$. For a rational consumer, the differential value delivered by the higher quality laptop 1 is:

$$\Delta u_{1,2} = (q_1 - q_2) - (p_1 - p_2) - \mu\left(f_1 - f_2\right). \tag{7}$$

Consider now a first time buyer with selective recall. While qualities and upfront prices are visible, maintenance costs are shrouded and tend not to come to mind, namely $a(\mu, 0) < a^*$. Moreover, as the consumer has no exposure to laptops he does not recall specific prices, $X_{p_k}(a^*) = \emptyset$, and the choice context is $\widetilde{C} = C_v = C_v(n) = \{(q_1, p_1), (q_2, p_2)\}$.

For simplicity, we restrict attention to the case in which in the visible choice set and thus in the context $C_v \cup C_v(n)$ quality is salient ($\frac{q_1}{p_1} > \frac{q_2}{p_2}$). As a result, FAST prefers the higher quality laptop 1 provided:

$$\Delta u_{1,2}^{FAST,q} = \frac{2}{1+\delta}\left(q_1 - q_2\right) - \frac{2\delta}{1+\delta}\left(p_1 - p_2\right) > 0. \tag{8}$$

Compared to the rational decision rule in (7), FAST is more likely to choose laptop 1 because: i) he neglects its higher maintenance costs, and ii) he overweighs its quality relative to its price. An FBOR overvalues laptop 1 only because of neglect of maintenance costs in i).

[17] In Equation (6) we assume for simplicity that laptops differ in their maintenance costs f_k, but the rate of maintenance μ is constant across laptops. In the case of other durable goods, what varies is the rate at which maintenance costs are incurred and not the per unit amount of these costs. For instance, in the case of a car, μ_k decreases in the car's miles per gallon while the cost of refilling the tank is constant, i.e. $f_k = p_{gallon}$, for all cars.

Formally, $\Delta u_{1,2}^{FBOR} = (q_1 - q_2) - (p_1 - p_2)$.

To assess the effect of information, suppose that the consumer is reminded about maintenance costs, μf_k for $k = 1, 2$. The consumer's choice and evoked sets then become comprehensive, $\widetilde{C} = C = C(n) = \{(q_1, p_1, f_1), (q_2, p_2, f_2)\}$. We have:

Proposition 1 *After information provision, the relative valuation for the high quality laptop by FBORs ($\delta = 1$) drops by the difference in maintenance costs, namely*

$$\Delta u_{1,2}^{FBOR}(q_k, p_k, f_k) - \Delta u_{1,2}^{FBOR}(q_k, p_k) = -\mu\,(f_1 - f_2) \leq 0.$$

For FASTs ($\delta < 1$), there are two thresholds $\overline{\Delta f} > \underline{\Delta f} > 1$ such that, relative to FBOR:

i) FAST underreacts to maintenance costs for $f_1/f_2 < \underline{\Delta f}$.

ii) FAST overreacts to maintenance costs for $f_1/f_2 > \overline{\Delta f}$.

When reminded of maintenance costs, an FBOR reduces his differential valuation of laptop 1 by the true utility difference $\mu\,(f_1 - f_2)$. FAST underreacts (overreacts) to maintenance costs when his differential valuation of laptop 1 drops less (more) than for FBOR. Proposition 1 identifies the conditions under which either takes place. When f_1/f_2 is close to one, maintenance costs are similar across laptops, and so are not salient. FAST's attention remains focused on laptop quality (and upfront prices), and he underreacts to the reminder. When instead f_1/f_2 is high, the laptops display large proportional differences in maintenance costs. In this case, these costs become salient and draw FASTs attention away from quality. As a consequence, FAST's valuation of laptop 1 drops too much relative to FBOR's.

As in Table 1, the interaction between memory and attention is critical here. The impact of information about the neglected attribute on behavior depends not only on whether that information goes against the ex ante choice of the expensive laptop 1, but also on whether the attribute that is now revealed is salient *relative* to the dimensions that were previously taken into account.

A striking illustration of overreaction to surprises is Hsee's (1998) dinnerware set experiment. Subjects must state their willingness to pay for a dinnerware set with 24 dishes in perfect condition and a dinnerware set with 31 similar dishes plus a few broken ones.

Subjects state a higher willingness to pay for the smaller dinnerware set, but choose the larger one in a direct comparison. The first result violates the basic economic assumption of monotonicity, i.e. more of a good is better (together with free disposal), while the two results together violate the assumption that choice preferences can be reconstructed by eliciting WTP (Tversky, Slovic and Kahneman 1990). Our model captures this finding as a combination of selective recall and surprise. When the consumer sees the larger dinnerware set, he is surprised to see the dimension of broken dishes (since a normal dish is in good condition). This draws attention away from the good dishes, lowering his WTP for the set. In direct comparison, however, it is evident for the consumer that the larger dinnerware set dominates. We work through this example formally in the Appendix.

5 Application 1: (Un)Shrouded Attributes

We now turn to a more detailed analysis of how the interaction between memory and attention can shed light on a variety of experimental and field evidence.

The literature on "shrouded attributes" stresses that firms often choose not to disclose all information relevant to the goods they sell, or to make this information difficult to access (Ellison 2005, Ellison and Ellison 2009). Sellers of printers shroud the cost of replacing a cartridge, hotels shroud the prices of add-ons such as mini-bar consumption or internet access.

Our model helps to think about these phenomena. First, the dependence of recall on exposure and frequency sheds light on which non-visible attributes consumers fail to recall. Second, selective attention allows us to analyze how consumers react to the unshrouding of hidden attributes. In Gabaix and Laibson (2006), unshrouding benefits unaware consumers, who are FBORs and thus optimally react to information. FASTs, in contrast, may over- or under-react to information. This logic helps explain a variety of real world phenomena, including the incentives for firms to unshroud certain product features.

Suppose a consumer is contemplating the purchase of a durable good, such as a TV, that is subject to the risk of a breakdown. In the notation of equation (1), total quality is given by $\theta_1 q + (1 - \theta_1) f$, where q is the utility of a functioning TV, $f = 0$ is the utility of a broken

TV, and $1-\theta_1$ is the (low) probability of breakdown. The price of TVs is deterministic and equal to p.

Consider first how the consumer represents the TV. Because TV breakdowns are rare, $\theta_1 \approx 1$, their availability in memory is low, and particularly so for consumers who have not had any exposure to breakdowns in the past, $e_f = 0$. In this case, the consumer may well neglect breakdown risks, formally $a(1-\theta_1, 0) < a^*$. In contrast, the working quality and the price of the TV are visible and top of mind, so the TV is represented as a quality price pair (q, p).[18]

Given the limited representation (q, p), the consumer's perceived choice set is $\widetilde{C} = \{(q, p), (0, 0)\}$. Since quality and price are deterministic, $\widetilde{C}$ is also the consumer's evoked set, $\widetilde{C}(n) = \widetilde{C}$. In the choice context $\widetilde{C} \cup \widetilde{C}(n)$, the salience of the TV attributes are measured relative to the reference quality $\overline{q} = q/2$ and the reference price $\overline{p} = p/2$. Due to homogeneity of the salience function, quality and price are equally salient, because $\sigma(q, \overline{q}) = \sigma(p, \overline{p}) = \sigma(1, 1/2)$. It follows that, when neglecting breakdown risk, FAST values the TV as $q - p$ and his willingness to pay for it is $WTP = q$. The first consequence of selective recall is straightforward: the consumer neglects breakdown risk so his willingness to pay is too high, by $(1-\theta_1)q$.

Nevertheless, sellers of TVs and home appliances regularly remind buyers about breakdown risks. They offer insurance against such events, which most buyers vastly overpay for (Sydnor 2010, Huysentruyt and Read 2010). The distinction between FBORs and FASTs is critical in thinking about this phenomenon.

5.1 Unshrouding with Insurance

After having chosen the TV, the consumer is informed by the salesperson of the risk that it may break down with probability $1-\theta_1$. This increases the consumer's exposure to breakdown risk, so that this possibility now comes to the consumer's mind, despite its low probability ($a(1-\theta_1, e_f) > a^*$). Upon unshrouding, the choice and evoked sets become

[18] Recent experience of a TV breakdown increases exposure and induces the consumer to think about this contingency, despite its low objective probability. Past personal experiences create heterogeneity in individual representations of choice situations, in line with the evidence of Malmendier and Nagel (2009). This is an empirically testable implication of our model.

comprehensive $\widetilde{C} = \widetilde{C}(n) = C = \{(q, p, 0), (0, 0, 0)\}$ and they both include breakdown risk as a normal quality contingency.

Consider first the "pure" effect of unshrouding, when insurance is not offered. In the new choice context, the reference TV attributes are $\overline{q} = q/2$, $\overline{f} = 0$, and $\overline{p} = p/2$. Homogeneity of the salience function implies $\sigma(q, \overline{q}) = \sigma(p, \overline{p}) = \sigma(1, 1/2) > \sigma(f, \overline{f})$, Quality and price are still equally (and maximally) salient for any price p. As a result, upon unshrouding, FAST's willingness to pay is $WTP = \theta_1 q_1$. Willingness to pay drops modestly, given that breakdown risk is unlikely. Even this modest drop, though, makes it unprofitable for the seller to unshroud.

In contrast, unshrouding may be profitable for the seller when coupled with insurance. Suppose that, when disclosing breakdown risk, the retailer offers the consumer an "enhanced" TV with insurance, namely a good (q, p', q) where $p' > p$. Now both the choice and the evoked set include the original uninsured TV, which is what the consumer had originally in mind, as well as the insured TV. The perceived choice and evoked sets are $\widetilde{C} = \widetilde{C}(n) = C_{ins} = \{(q, p', q), (q, p, 0), (0, 0, 0)\}$. In this choice context, the risk of breakdown is very salient, because there is a sharp contrast between the attributes delivered by the insured TV versus the uninsured TV in that contingency. Specifically, the reference attributes are $\overline{q} = 2q/3$, $\overline{f} = q/3$, and $\overline{p} = (p' + p)/3$. Homogeneity of the salience function implies that breakdown risk is the most salient attribute of both the insured and the uninsured TV.[19] Breakdown risk is salient because it is the attribute along which the insured and the uninsured TV are most different. Crucially, the salience of breakdown risk increases the perceived value of insurance to the point that the consumer may be willing to "overpay" for the TV. Formally:

Proposition 2 *Let WTP denote the FAST consumer's willingness to pay for the insured TV in C_{ins}. If δ is sufficiently low, then:*

i) $WTP > p + (1 - \theta_1)q$, i.e. the consumer is willing to overpay for insurance.

[19]Formally:

$$\begin{aligned} \sigma\left(q, \frac{q}{3}\right) &= \sigma\left(1, \frac{1}{3}\right) > \max\left[\sigma\left(p', \frac{p'+p}{3}\right), \sigma\left(q, \frac{2q}{3}\right)\right], \\ \sigma\left(0, \frac{q}{3}\right) &> \max\left[\sigma\left(p, \frac{p'+p}{3}\right), \sigma\left(q, \frac{2q}{3}\right)\right] \end{aligned}$$

where the top row refers to the insured TV while the bottom row refers to the insured TV.

ii) if $p = q$ and the seller pays the actuarially fair price for insurance, the seller's profit is higher when he sells the insured TV than when breakdown risk is shrouded.

Because breakdown risk is salient, the consumer overweighs the benefit of insurance, and is willing to pay an implicit premium for it that is above the actuarially fair price, particularly if δ is low.[20] It is then more profitable for the seller to remind consumers of the breakdown risk and sell the overpriced insurance premium than to sell overpriced TVs to consumers who neglect breakdown risk.[21]

5.2 Memory, Attention, and Shrouded Attributes

This example shows that the interaction between selective memory and selective attention leads to new implications of unshrouding. Again, both ingredients play a critical role.

Consider the role of selective attention first. When consumers are FBORs, $\delta = 1$, sellers have no incentive to unshroud breakdown risk, even when simultaneously offering insurance. Indeed, the value of the insurance for an FBOR who learns about the breakdown risk is exactly $(1 - \theta_1)\, q$, so the retailer cannot sell an insured TV for more than the willingness to pay $p = q$ of a forgetful consumer. In contrast, sellers do have an incentive to unshroud breakdown risk when consumers are FAST, because FASTs over-react to the breakdown contingency. Because of this over-reaction, unshrouding plus insurance allows the retailer to charge higher prices.

Selective memory also plays a key role in this analysis. First, it is required for the breakdown risk to be neglected in the first place. It also shapes over-reaction to unshrouding, for it leads the consumer to evaluate the insurance option in light of the fallback $(q, p, 0)$ of no insurance. If the consumer recalled the breakdown risk and thought about buying an insured TV from the start, he would value insurance correctly. In this case, $C = C(n) =$

[20]The literature has linked insurance choices, and in particular overpaying for insurance against small risks such as appliance breakdown risk, to overweighting of small probabilities and more generally to the properties of the probability weighting function of Prospect Theory (Kahneman and Tversky 1979, Sydnor 2010). In BGS (2012) we argue that salient thinking provides a foundation for such distortions, and describe the set of circumstances in which they arise.

[21]Along the same lines, the true market value of insurance is also misrepresented, as it is measured against the value q of the existing TV rather than against the market price p^n of getting a new TV. To the extent that $p < q$, this is also a source of distortions.

$\{(q, p', 0), (0, 0, 0)\}$, and since all attributes are equally salient, the consumer's valuation is rational.[22] FAST pays too much precisely because he forgets that a normal TV can break. When this risk is brought up, the value of insurance is salient.[23]

In sum, the combination of selective memory and selective attention can explain both why consumer can neglect product attributes and why it can be profitable for firms to inform them. As Proposition 2 (point ii) shows, in our model informed FAST consumers are not necessarily better off, and unshrouding may benefit firms to their disadvantage. In Gabaix and Laibson (2006), in contrast, all consumers are FBORs and in the market equilibrium the uninformed cross subsidize the informed. Here we do not solve for market equilibrium, but the example above suggests that when consumers are FAST, cross subsidization may work the other way around.

This logic can also provide insight into recent findings on consumer demand for usage plans with non-linear pricing: Grubb (2009, 2014) shows that consumers on average choose cell-phone plans that include too large usage allowances relative to their needs. Ater and Landsman (2013) document the same pattern for retail banking services. These findings suggest a high willingness to pay for insurance, consistent with Proposition 2. Moreover, Ater and Landsman (2013) show that consumers that happen to go beyond their allowance, and therefore experience overage fees, are more likely to switch to plans with larger allowances. The authors show that these consumers avoid overage fees but end up paying more on average than before. This seems consistent with an overreaction to the experience of overage fees.[24]

[22] Alternatively, if a TV costs p, the consumer who recalls the background risk may factor in his initial choice the expected cost $(1-\theta_1)\,p$ of replacing a broken TV, i.e. he considers the option $(q, q, p + (1-\theta_1)\,p)$. When comparing this option to the insurance offered by the retailer (q, q, p'), then, price is salient. Intuitively, price is the only attribute that varies across choice options. As a result, the consumer only buys the insured TV if $p' \leq p + (1-\theta_1)\,p$, namely if the cost of insurance does not exceed the expected cost of buying a new TV.

[23] As we show in BGS (2013), salience distortions arise when goods are differentiated, even in the absence of surprising attributes or price realizations. However, salience plays a further role when recall is selective: unshrouding an attribute can be surprising if it changes the ex post allocation of attention. While in this section we focus on shrouded attributes, this logic provides insight into why and how add-ons more generally can be profitable (Ellison 2005): when a consumer has in mind a good he plans to buy, firms can offer an add-on whose upside is salient when compared to that good.

[24] This also suggests a rationale for why firms resist "bill-shock" regulation. Firms make a significant part of their revenue on such overage fees (Grubb 2009, 2014). How would a consumer react to being informed that he has reached the limit of his allowance? Intuitively, the normal marginal price for a phone call is zero (since that is the price the consumer is most exposed to). However, when the allowance is spent, the consumer suddenly faces a much higher marginal price, the overage fee. In this case, consumers may overreact

More generally, the interaction between selective memory and attention provides some insight into a longstanding tension in the literature between evidence suggesting that people neglect low likelihood contingencies and evidence suggesting such contingencies are overweighted (Kahneman and Tversky 1979). Low likelihood contingencies tend to be neglected because they have low frequency and get low exposure. Yet, once people are reminded of them, they can become very salient (Browne and Hoyt 2000).

6 Application 2: Effects of Reminders

A large body of empirical work in applied microeconomics seeks to assess the extent to which suboptimal decisions can be accounted for by selective recall. To quantify the role of selective recall, these studies measure changes in behavior occurring when "shrouded" aspects of the choice problem are made more visible to consumers. Chetty et al (2009) documents that demand for certain products falls when consumers are reminded of sales taxes (see also Goldin and Homonoff, 2013). Finkelstein (2009) shows that highway usage becomes more price sensitive when tolls are more visible (see also Sexton 2014). Reminders change behavior in the expected direction in many other settings: they help increase exercise (Calzolari and Nardotto, 2014) and savings (Karlan et al, 2014), they help reduce weight (Patrick et al. 2009) and overdraft fees (Stango and Zinman, 2011).

The conventional view on the effects of reminders rests on two pillars: 1) consumers forget about certain relevant attributes of the choice problem, and 2) the change in behavior reflects the optimal use by consumers of information contained in the reminders. In our model, this interpretation amounts to assuming that a reminder about a neglected attribute j increases the consumer's exposure e_j to it. Consumers are assumed to be FBORs, so that the change in their behavior corresponds exactly to the elimination of neglect. This perspective has important welfare implications: in a world of FBORs, changes in behavior due to information provision offer a proxy for the welfare costs of forgetting, providing guidance on the role of "nudges" in public policy (Bernheim and Rangel 2009).

to this salient price hike by reducing their consumption over and beyond what is justified by the increased marginal cost. In this way, reminders of costs may significantly reduce firms' revenue from overages. This is also consistent with Stango and Zinman's (2011) findings in the context of overdraft fees.

Our analysis shows that the psychology of reminders is more complex. In most circumstances, increased exposure e_j does more than to bring attribute j to top of mind, providing the consumer with additional data. By shaping attribute salience, it also affects the attention allocated to data that were *already available* to the consumer. The effect of reminders on behavior conflates these two aspects. In this sense, the measured effect of reminders cannot a priori be considered an accurate estimate of the cost of forgetting.

This section illustrates this idea by revisiting several recent experiments in light of our model. We identify the conditions under which consumers may under- or overreact to reminders. From a positive standpoint, our analysis can help explain why the strength of the reminder effects is found to vary: for example, when people are reminded to save, they indeed save more, but reminding people of their future goals leads to much higher savings rates (Karlan et al, 2014). These differences can be understood as reflecting variations in the salience of reminders. From a normative standpoint, our model offers guidance on how to unpack the effects of reminders on memory and attention, and on how to design effective reminders.

6.1 Tax Salience

Chetty et al. (2009) report an experiment in which posting prices inclusive of sales taxes for some goods reduced consumer demand for them, relative to the case in which only pre-tax prices were posted. They account for this finding by arguing that sales taxes paid at the cashier are less "salient" than posted prices, and may be forgotten by shoppers. When reminded of the sales tax, shoppers realize that the good is more expensive than they thought and optimally reduce their demand.

To support this interpretation, Chetty et al (2009) present survey evidence that i) given pre-tax prices, the surveyed population (students) typically neglects taxes when computing expenditures, but accurately incorporates taxes when reminded of them; and ii) consumers at the store have accurate knowledge of the sales tax levied on a variety of goods. In this account consumers are FBORs who react rationally to information about a neglected attribute. In line with this interpretation, Chetty et al (2009) propose a methodology to infer consumers' true preferences from their behavior under tax-inclusive prices.

Our model suggests that matters may be more complex. Suppose a consumer is deciding whether or not to buy a good (q, p, τ) consisting of quality, price, and sales tax (τ represents the dollar value of the sales tax). All attributes are deterministic. As stressed in the literature, the sales tax is forgotten because it is added on to the bill at the cashier and consumers generally do not associate specific tax amounts with individual goods. In our model, this idea is captured through exposure: even though the tax is paid with probability one, consumers' exposure to it is sufficiently low $a(1, e_\tau) < a^*$ that the tax does not come to mind.[25]

Because he neglects the sales tax, the consumer represents good (q, p, τ) as (q, p), and perceives the choice set, as well as the evoked set, to be $\widetilde{C} = \widetilde{C}(n) = C_v = \{(q, p), (0, 0)\}$. In this choice context, $\overline{q} = q/2$ and $\overline{p} = p/2$, so that quality and price are equally salient, $\sigma(q, \overline{q}) = \sigma(p, \overline{p}) = \sigma(1, 1/2)$. Because quality and price are equally salient, FAST's valuation of the visible attributes is undistorted. As a consequence, both FBOR and FAST value the good at $q - p$. This would be the rational valuation if the sales tax was zero.

Suppose now that, as in Chetty et al (2009)'s intervention, the posted price is tax inclusive, so the price of the good is $p' = p + \tau$. How does this change affect FAST's valuation? There are are two ways in which the consumer may encode the intervention, which have different implications for choice behavior. We now discuss each of these perspectives.

Perspective 1: Tax Unshrouding. From this perspective, when the consumer sees price $p' = p + \tau$, he fully realizes that it is due to the inclusion of the sales tax. This brings the tax dimension to mind, restoring the three dimensional representation of the good as (q, p, τ). The consumer's choice and evoked set become $\widetilde{C} = \widetilde{C}(n) = C = \{(q, p, \tau), (0, 0, 0)\}$. In this choice context, $\overline{q} = q/2$, $\overline{p} = p/2$ and $\overline{\tau} = \tau/2$ so quality, price and the tax are (again) equally salient, because $\sigma(q, \overline{q}) = \sigma(p, \overline{p}) = \sigma(\tau, \overline{\tau}) = \sigma(1, 1/2)$. As a result, the consumer's perceived utility is still non-distorted.

Formally, after the reminder the consumer's utility from the good is $q - p - \tau$, and the consumer buys the good provided $q \geq p + \tau$. After a reminder about taxes, the consumer's

[25] To repeat the terminology discussed in the Introduction, we refer to information that comes to mind as recalled (or forgotten), and information that draw a disproportionate share of attention, conditional on what is remembered or observed, as salient.

valuation of the good drops. The drop in valuation, though, is "rational", in the sense that the new value of the good is the fully rational re-optimised decision rule. This is intuitive: bringing the tax to mind leads to a revision of the "norm" for the good, which now includes taxes. Because the level of the tax is itself normal, the consumer is not surprised and perfectly integrates the tax into his decision. From this tax-unshrouding perspective, FAST's behavior is thus consistent with Chetty et al (2009)'s interpretation of consumers as FBORs.[26]

Perspective 2: Price Surprise. FASTs react to reminders like FBORs only in special circumstances. In Chetty et al.'s experiment, this depends on how the consumer interprets the experimental treatment, namely the posting of the tax inclusive price.

To see this, suppose that when the consumer sees price $p' = p + \tau$, he fails to realize that it is due to the inclusion of the sales tax. That is, the consumer continues to neglect the tax and instead thinks that the good has become more expensive. For an FBOR, this misunderstanding does not matter: given that price and tax are fungible, i.e. equally weighted, a given increase in cost due to either base price or tax leads to the same drop in valuation.

For FAST, the equivalence between price and tax breaks down because of salience. Faced with the new higher price, FAST retains a two dimensional representation of the good, with a choice set $\widetilde{C} = C_v = \{(q, p'), (0, 0)\}$. Crucially, the corresponding evoked set is $\widetilde{C}(n) = \{(q, p), (0, 0)\}$, namely the visible choice set at the normal "tax exclusive" price. In this choice context, the reference attributes are $\overline{q} = q/2$ and $\overline{p} = (p + p')/4$. Thus, under the price surprise perspective, the reminder causes price to be more salient than quality ($\sigma(p', \overline{p}) = \sigma(1, (p+p')/4p') > \sigma(q, \overline{q}) = \sigma(1, 1/2)$ because $p' > p$). The consumer's valuation then drops from $q - p$ to:

$$\frac{2}{1+\delta}\left[\delta q - p(1+\tau)\right].$$

[26]This result does not rely on the fact that consumers are choosing between (q, p, τ) and $(0, 0, 0)$. It is also consistent with the possibility that consumers may be choosing between goods subject to the same sales tax, (q_1, p_1, τ_1) and (q_2, p_2, τ_2). In this case, salience may distort utility and choice if quality and price are not equally salient (i.e., if $\sigma(q_k, \overline{q}) \neq \sigma(p_k, \overline{p})$). However, the tax would always be as salient as price, for $\sigma(\tau_k, \overline{\tau}) = \sigma(\tau p_k, \tau \overline{p}) = \sigma(p_k, \overline{p})$. Hence, the consumer would always view the tax as being fungible with the price, just like a rational consumer.

In contrast, even from the tax unshrouding perspective the tax would look surprising if the good in question is compared to goods that are not subject to the sales tax. These cross-good effects are a testable implication of our model.

We therefore have:

Proposition 3 *From the "price surprise" perspective, FAST overreacts to the tax inclusive price relative to the "tax unshrouding" benchmark. The excess reduction in valuation in this case is* $\left(\frac{1-\delta}{1+\delta}\right)[q+p+\tau]$.

When consumers view the tax inclusive price as a price increase, they are surprised by what they perceive as higher than normal prices. As a consequence, prices become salient and valuation drops by more than the price hike. The consumer over-reacts to information relative to the case in which he understands that the price hike is due to sales taxes.

To interpret experiments such as Chetty et al (2009) it is important to pin down the way in which consumers interpret information. Chetty et al (2009) make serious efforts to ensure that consumers understood that the higher price was the tax-inclusive price, but they do not present direct evidence that this is the case. While this distinction is immaterial if consumers are FBOR, it is crucial if consumers are FAST. Proposition 3 implies that the possibility that (some) FAST consumers mistook the intervention for a price increase has important implications for the interpretation of its effects. More generally, if consumers are FAST, then the mechanism of overreaction of Propositions 1 and 3 is relevant for welfare analysis.

In the Chetty et al. experiment, our analysis suggests the following hypothesis: the clearer it is to consumers that the higher prices simply reflect the sales tax, the smaller the drop in demand should be (converging towards the rational consumer case).[27]

Generally speaking, this analysis suggests that experiments relying on changes in the visibility of price/cost variation over time (Finkelstein 2009, Sexton 2014) or across goods (Allcott and Taubinsky 2014) are likely to combine information provision with salience effects. When intertemporal or cross sectional price differences become more visible, two forces operate in the consumers' mind. First, consumers are provided with more accurate information. Second, the new information may either be striking enough to distract from existing

[27]An experiment to test this hypothesis would have two treatments: in one treatment, only tax-inclusive prices are presented, with a minor reference to the tax being included ("Full Price, Inclusive of Sales Tax"). In the other treatment, it is made clearer to the consumer that the higher prices simply reflect the sales tax ("Our prices have not changed, but for your convenience we now include the full tax, inclusive of sales taxes"). A difference in demand between the treatments would indicate a channel other than the fact that the treatments themselves draw attention to prices.

information on quality, or it may be unsurprising enough to be downplayed relative to quality differences. In these cases, the salience mechanism generates over-reaction or under-reaction to information, acting as a confound for the pure informational role of reminders.

6.2 Reminders and Savings

Karlan et al. (2014) explore the effects of reminders on saving rates when consumers have committed to save for a goal (e.g., insurance against job loss). They show that reminding people of their goals leads to a large increase in savings relative to what is obtained under a generic reminder to save.

We now show how our model accounts for this finding as the result of the variation in salience across reminders. A risk neutral consumer lives for two periods and is indifferent between consumption in the two dates. In period 1 he receives an endowment y. In period 2 his endowment is given by:

$$y_2 = \begin{cases} y & \text{with probability} \quad \theta \\ 0 & \text{otherwise} \end{cases},$$

where θ is close to one. The zero endowment state captures an unlikely bad event in which, for example, the consumer becomes sick and loses his labor income.

The consumer must choose whether to save a given amount $s^* < y$ in period 1. Savings are deposited in a bank account that pays an interest rate $r > 0$. Given savings $s \in \{0, s^*\}$, consumption in the final period $t = 2$ is equal to:

$$c_2 = \begin{cases} y + s(1+r) & \text{with probability} \quad \theta \\ s(1+r) & \text{otherwise} \end{cases}.$$

The consumer's indifference between consumption at different dates implies that a rational consumer optimally saves $s = s^*$ to earn the interest income rs^*.

Consider now the behavior of a FAST. A consumption-savings plan consists of a three dimensional vector $\left(c_1, c_2^h, c_2^l\right)$ reporting consumption at different dates and states (where h and l respectively represent the good and bad states at $t = 2$). In line with Karlan et al., we assume that in the absence of reminders the consumer neglects future consumption. Intuitively, while exposure to present consumption is high, cues about future consumption are more sporadic, and so is exposure. Formally,

$$a(1, e_{c_1}) > \overline{a} > a(\theta, e_{c_2^h}) > a(1 - \theta, e_{c_2^l}).$$

In the absence of reminders, the consumer neglects the future states, maximizes his current consumption c_1, and chooses not to save at all.[28]

Consider now the effects of two particular reminders. The first reminder generically encourages people to save "for the future". The second reminder is targeted, and encourages people to save "for an emergency".

Consider the generic reminder first. The message "save for the future" enhances exposure to all future states, increasing both $e_{c_2^h}$ and $e_{c_2^l}$. However, because the bad state has a low objective probability, consumers are likely to continue to neglect it. Formally, we assume $e'_{c_2^h} > e_{c_2^h}$, $e'_{c_2^l} > e_{c_2^l}$ and

$$a(1, e_{c_1}) > a(\theta, e'_{c_2^h}) > \overline{a} > a(1 - \theta, e'_{c_2^l}).$$

The generic reminder brings to mind the more likely future state h, but not the unlikely disaster state l. Thus, the consumer thinks about his consumption savings problem as choosing a two dimensional vector (c_1, c_2^h), defined by the perceived choice set:

$$\widetilde{C} = \left\{(y, y), (y - s^*, y + s^*(1 + r))\right\}.$$

The evoked set satisfies $\widetilde{C}(n) = \widetilde{C}$. In this choice context, the first period consumption

[28] Note that this phenomenon is different from impatience. Under impatience, the consumer would rather borrow in order to increase current consumption. In this case, neglect of the future implies that borrowing does not even come to mind.

(and hence the cost of saving) is salient provided:

$$\frac{s^*}{y} > \frac{r}{1+r}. \tag{9}$$

When the interest rate r is low relative to the target savings rate s^*/y, FAST overweights present (loss of) consumption relative to future consumption. As a consequence, when Equation (9) holds, FAST behaves impatiently provided $(1+r)\,\delta < 1$. Combining the two conditions, we find that for

$$r < \overline{r} = \min\left\{\frac{1}{\delta} - 1, \frac{s^*}{y - s^*}\right\},$$

the generic reminder is ineffective in persuading FAST to save, for the benefits of saving are less salient than its costs. Instead, an FBOR realizes that savings is optimal due to the positive interest rate; the reminder causes him to save even if state l does not come to mind.

Suppose now that the consumer is specifically reminded to save for future emergencies. This reminder increases exposure to the future, but particularly to the bad state l. Formally, we assume that the targeted reminder brings to the consumer's mind not only the likely good state h but also the disaster state l. In this case, the consumer compares the full, three dimensional, consumption streams. For FAST, the most salient state is always the disaster state, in which not saving yields a very salient income of zero. If (9) holds, FAST saves provided the interest rate is high enough:

$$r > \underline{r} = \frac{1}{\theta_L\left(\frac{1}{\delta} - \delta\right) + \delta} - 1.$$

Because $\underline{r} < \overline{r}$, there is a range of interest rates $r \in [\underline{r}, \overline{r}]$ for which FAST does not save under the generic reminder but does save under the targeted reminder. Salience is key in accounting for the different effects of the reminders: the targeted reminder is more effective because it makes people think of the state in which savings make the largest difference, and thus are salient.

Karlan et al (2014) assume, as we did above, that the generic reminder might fail to bring their savings goals to the subjects' minds. In our model, this occurs because savings goals are often related to expenses that are infrequent, such as school fees, or unpredictable, such as “an emergency”. These expenses have lower availability than more frequent, or more immediate, ones (Sussman and Alter, 2013). To obtain strong savings responses, it is more effective to explicitly remind subjects of their goals. Unlike in Karlan et al (2014), however, our model suggests that doing so may cause savings to become too salient, and subjects to save too much.[29]

To conclude, generic reminders may not have the desired effect if they do not bring to mind contingencies with salient consequences. For instance, reminding people to eat healthy, or to save, may be ineffective if it fails to bring to mind the drastic implications of a given choice for longterm health, or financial security. This prediction of our model is related to Akerlof (1991). The form and content of the reminder may have large implications on its effectiveness, depending on what it brings to mind.

6.3 Empirical Estimates of Reminder Effects

We conclude this section by showing the implications of our model for the interpretation of the effects of reminders. To do so, we augment the standard analysis of the demand response to reminders with the possibility that consumers may be FASTs.

There is a measure one of consumers, each of whom chooses between two goods $k = 1, 2$. The true utility of consumer $i \in [0, 1]$ from good (q_k, p_k, f_k) is equal to:

$$u_{i,k} = \theta_i q_k - p_k - f_k,$$

where $\theta_i > 0$ captures a consumer-specific taste for quality, distributed in $[0, +\infty)$ according to c.d.f. $F(\theta_i)$.

Good 1 has higher (visible) quality and upfront price than good 2, namely $q_1 > q_2$ and $p_1 > p_2$. Attribute f_k is a shrouded cost, which is also higher for good 1, namely $f_1 > f_2$.

[29] In our approach, the different attributes of choice capture consumption in different time periods (as in Koszegi and Szeidl 2013) and in different states (as in BGS 2012). As a consequence, consumers exhibit impatience if earlier periods are more salient than later ones, but are too patient if the reverse salience ranking holds. This instability of time preferences (among the periods that come to mind) is related to that in Koszegi and Szeidl (2013). See Bushong, Rabin and Schwartzstein (2014) for an alternative mapping of consumption streams into sets of attributes.

In Chetty et al (2009), f_k is a sales tax; in Alcott and Taubinsky (2014), f_k is a lightbulb's energy consumption.[30]

We consider the behavior of FASTs in the simplest case in which quality and upfront price are equally salient, namely $\sigma(q_k, \overline{q}) = \sigma(p_k, \overline{p}) = \widetilde{\sigma}$ for $k = 1, 2$.[31] The analysis could be more general, but it is convenient to restrict salience distortions to affect only the tradeoff between "visible" utility carriers $(\theta_i q_k - p_k)$ and the shrouded cost f_k. In this case, FBORs and FASTs differ only in their reaction to the reminder of shrouded costs. The utility *perceived* by FASTs who are informed about the shrouded attribute f_k is then:

$$u^S_{i,k} \propto \theta_i q_k - p_k - \phi f_k, \tag{10}$$

where:

$$\phi = \begin{cases} \delta & \text{if} \quad \sigma\left(f_k, \overline{f}\right) < \widetilde{\sigma} \\ 1 & \text{if} \quad \sigma\left(f_k, \overline{f}\right) = \widetilde{\sigma} \\ 1/\delta & \text{if} \quad \sigma\left(f_k, \overline{f}\right) > \widetilde{\sigma} \end{cases} . \tag{11}$$

The informed consumer underweighs the shrouded cost $(\phi = \delta < 1)$ if the latter is less salient than quality and price, while he overweighs it $(\phi = 1/\delta > 1)$ if the reverse is true.

Consumers are heterogeneous in what they recall. A share α of consumers remembers the shrouded attribute, while the remaining share $(1 - \alpha)$ forgets it.[32] Remembering FASTs value good k according to Equation (10). Forgetting FASTs, in contrast, represent good k

[30] To map this setting to Alcott and Taubinsky (2014), we assume that good 1 is the incandescent light bulb, which has a nicer color of light (higher q) but also higher expected upfront price (due to a higher replacement rate) and higher energy consumption requirements f, than the more efficient CFL model.

[31] One reasonable condition for equal salience occurs when quality is produced under constant returns to scale (i.e. the cost of quality is $C(q) = q$) and price is linear in cost, namely $p(q) = bC(q)$ for some $b > 0$. This property is due to the homogeneity of degree zero of the salience function.

[32] Whether a consumer is informed or not is independent of the consumer's taste θ_i. One can think about consumer heterogeneity as stemming from consumer-specific exposure $e_{i,f}$. These differences in exposure induce differences in availability $a_{i,f}$. If the availability of shrouded costs is distributed according to c.d.f. $A(a_{i,f})$, then we have that:

$$\alpha = 1 - A(\overline{a}).$$

In this setting, a reminder increases exposure for all consumers, increasing the mass of consumers whose availability is above $\overline{a}$. This intervention, then, reduces $A(\overline{a})$ thereby increasing the share of aware households α.

as (q_k, p_k) and value it as $\theta_i q_k - p_k$.

Denote by Δx the difference $(x_1 - x_2)$ in the value of attribute $x = q, p, f$ across goods. The total demand for the cheaper good 2 is given by:

$$D = \alpha F\left(\frac{\Delta p + \phi \Delta f}{\Delta q}\right) + (1-\alpha) F\left(\frac{\Delta p}{\Delta q}\right), \tag{12}$$

which includes a share of remembering consumers (first term on the right), and a share of forgetting consumers (second term). Relative to the rational benchmark, forgetting consumers neglect the shrouded cost Δf. Remembering consumers take Δf into account, but they over- or under-weigh it by the factor ϕ.

A reminder here can be viewed as an intervention that increases the share of remembering consumers from α to $\alpha' > \alpha$. Under a linear approximation of the difference between the demand of the two types of consumers, the reminder causes the demand for good 2 to change as:

$$\frac{D' - D}{D} \approx (\alpha' - \alpha) |\eta| \phi \frac{\Delta f}{\Delta p}. \tag{13}$$

where $|\eta|$ is the (absolute value of the) elasticity of demand of FBOR consumers.

The effect of the reminder on the demand for good 2 combines two effects. First, there is rational response to the reminder. As long as $\Delta f > 0$, an FBOR population would react to the reminder by increasing demand for the cheaper good by $(\alpha' - \alpha) |\eta| \frac{\Delta f}{\Delta p}$. The second effect is the change in salience triggered by the reminder itself, which is captured by the factor ϕ. If the unshrouded cost is salient ($\phi > 1$), informed consumers overweight it. The change in demand is then too large relative to the informational content of the reminder, and reflects consumers' over-reaction. If instead the shrouded cost is not salient ($\phi < 1$), informed consumers underweight it. The change in demand is then too small, reflecting under-reaction to information.

When $\phi = 1$, namely when FASTs behave as FBORs, Equation (13) recovers the formula used by Chetty et al. (2009) to quantify the welfare cost of selective recall from observed demand changes. In the Chetty et al (2009) experiment, the reminder about sales tax does not give rise to attention distortions if consumers have correct expectations about taxes. (This requires subjects to correctly interpret the higher price as being tax inclusive, which the experiment tries hard to do). Because $\phi = 1$, the demand response reflects the cost of forgetting.

In other instances, however, the presence of a salience distortion ϕ makes it more difficult to interpret the change in demand. In these cases, the interaction between memory and attention can help interpret existing evidence. Consider for instance the Alcott and Taubinsky (2014) experiment, in which consumers are informed of the energy costs of using incandescent and fluorescent lightbulbs. In this experiment, the presence of salience effects seem plausible. On the one hand, the information on energy costs given to the subjects appears to be render such costs more salient than upfront prices, since they display a larger percentage variation. On the other hand, quality differences across lightbulb technology also seems salient (i.e. CFLs have higher disposal cost and lower performance than incandescent lightbulbs). This creates a potential challenge for empirical identification. If in fact consumers focus on quality differences, and under-react to information on energy cost savings, then the small demand response to the reminder does not imply that the benefit of fluorescent bulbs is low. Indeed, a low demand response may simply be a result of under-reaction, and the paper's conclusion might be reversed, in that a ban on incandescent lightbulbs may be welfare increasing.[33,34]

In the preceding discussion, we implicitly assumed that FAST's over and under-reaction to the reminder (namely, whether the shrouded attribute is surprising) is normatively irrelevant. In doing so, we followed the literature (Bernheim and Rangel 2009, Mullainathan, Schwartzstein and Congdon 2012) in assuming that only the *informational* effect on demand matters. This assumption has two important implications in the context of our model:

- The model provides insight into the design of experiments that are robust to salience effects. The cost of selective memory on welfare is best assessed in experiments that elicit consumers' willingness to pay – namely, that assess demand relative to an outside

[33] Allcott and Taubinsky (2014) document that this is the case if there are sufficiently strong behavioral biases that are not corrected by information provision.

[34] In a similar vein, consider Abaluck (2011)'s study of the impact of nutrition labels on food consumption. He finds that on average people do reduce consumption of fat, but by a reduced amount that corresponds to a very small health improvement. Abaluck argues that this small reaction to nutritional information is difficult to reconcile with the much larger value of a healthy diet, as measured by value of the corresponding extension in life expectancy. Our model suggests that people under-react to the nutritional labels because, at the moment of choice, that information is less salient than other dimensions such as taste or even calorie intake (see also Koszegi, Szeidl 2013).

option of not buying – when consumers expectations about the information conveyed by the reminder are on average correct. This avoids surprises relative to expectations (and salience distortions relative to alternatives of choice), so that the consumer's reaction to information is rational. The sales tax experiment in Chetty et al (2009) is an example of such a setting. In Allcott and Taubinsky (2014), this might be achieved by eliciting WTP of lightbulbs separately.

- The literature suggests a second strategy to extract welfare implications from a demand response (13) when a behavioral model is given, namely to estimate the parameters of the model in an appropriate set of experiments and then use these parameters to decompose the demand response to the reminder into informational and behavioral effects, see Mullainathan et al (2012) and Chetty (2015). In our model, behavioral biases – i.e. the salience ranking of attributes – are context-specific. The salience ranking generated by one set of experiments may not coincide with that generated by the reminder intervention. This "Lucas critique" logic implies that such an estimation approach is infeasible. Instead, our model suggests that to correctly interpret a reminder experiment, detailed knowledge of the informational structure accessible to consumers is necessary. In different contexts, different biases (e.g. under- or over-weighting costs) may emerge.

7 Conclusion

We have presented a model of intuitive choice in which both memory and attention are limited and selective. We have stressed the idea that the recall process influences what is salient in the choice set, draws attention to those attributes, and therefore makes them more influential in decisions. This formulation naturally leads to a definition of surprise and its influence on choice, and helps account for a variety of evidence. In particular, it suggests that reaction to information need not be rational: decision makers can overreact or underreact depending on whether they are surprised by or oblivious to information. We applied this idea to account for, in a unified way, empirically documented findings ranging from rental choices, to shrouded attributes, to overage charges, to reaction to reminders.

More generally, our analysis suggests a new approach to modeling reference points that combines backward looking and forward looking elements. In our analysis, recall from memory is a fundamental feature of reference point formation, and is shaped in part by personal exposure, and in part by objective reality. This approach is more flexible than either a pure backward looking perspective (where the reference point is given by history or the status quo) or the pure forward looking view, where reference points are shaped entirely by expectations. Nonetheless, we have argued that this approach helps account for a great deal of evidence, but also has strong empirical implications working through exposure or other sources of changes in the availability of information.

Some of the most interesting implications of our analysis are barely touched on here, but suggest some avenues for future work. In particular, our model points to a broader theory of advertising and persuasion. At the most basic level, since many goods require a certain threshold of exposure to come to the top of mind, the model suggests that one purpose of advertising and persuasion is simply to increase exposure so that attributes (or goods) are recalled when the consumer needs to make decisions. But there is a more subtle, but perhaps equally important, implication of the analysis. Communication of information about attributes of a product automatically brings to mind these same products with normal values of these attributes. Persuasion changes what the consumer remembers. It follows that effective advertising must bring out the attributes that are sufficiently attractive that they can become salient relative to what is normal, and therefore draw attention and increase demand. In this regard, advertising of normal attributes will not be nearly as effective. The purpose of persuasion more generally to bring out aspects that are both attractive to the listener and distinctive. It is not what the persuader says directly, but also what the message brings to mind, that might be at the heart of which messages work.

References

Akerlof, George, 1991. "Procrastination and Obedience." *The American Economic Review, Papers & Proceedings* 81 (2): 1 – 19.

Allcott, Hunt and Dmitry Taubinsky. 2015. "Evaluating Behaviorally-Motivated Policy: Experimental Evidence from the Lightbulb Market" *American Economic Review*, forthcoming.

Ariely, Dan, George Loewenstein and Drazen Prelec. 2003. "Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences." *Quarterly Journal of Economics* 118(1): 73 – 105.

Ater, Itai and Vardit Landsman. 2013. "Do Customers Learn from Experience? Evidence from Retail Banking." *Management Science* 59 (9): 2019 – 2035.

Atkinson, Richard and Richard Shiffrin. 1969. "Storage and Retrieval Processes in Long-Term Memory." *Psychological Review* 76(2): 179 – 193.

Baucells, Manel, Martin Weber, Frank Welfens. 2011. "Reference-Point Formation and Updating." *Management Science* 57(3): 506 – 519.

Bernheim, Douglas and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics" *The Quarterly Journal of Economics* 124 (1): 51 – 104.

Bordalo, Pedro, Nicola Gennaioli and Andrei Shleifer. 2012. "Salience Theory of Choice under Risk." *Quarterly Journal of Economics* 127 (3): 1243 – 1285.

Bordalo, Pedro, Nicola Gennaioli and Andrei Shleifer. 2013. "Salience and Consumer Choice." *Journal of Political Economy* 121(5): 803 – 843.

Bordalo, Pedro, Nicola Gennaioli and Andrei Shleifer. 2014. "Stereotypes." Mimeo, Harvard University.

Bodner, Ronit and Drazen Prelec. 1994. "The Centroid Model of Context Dependent Choice", Unpublished Manuscript, MIT.

Browne, Mark and Robert Hoyt. 2000. "The Demand for Flood Insurance: Empirical Evidence." *Journal of Risk and Uncertainty* 20(3): 291 – 306.

Bushong, Benjamin, Mathew Rabin and Joshua Schwartzstein. 2014. "A Model of Relative Thinking", Mimeo Dartmouth College.

Calzolari, Giacomo and Mattia Nardotto 2014. "Nudging with Information: a Randomized Field Experiment." CEPR Discussion Paper No. DP8571

Chetty, Raj, Adam Looney and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence", *American Economic Review*, 99(4): 1145-1177.

Chetty, Raj. 2015. "Behavioral Economics and Public Policy: A Pragmatic Perspective" *American Economic Review Papers and Proceedings* (Ely Lecture), forthcoming.

Consumer Reports. 2005. "Extended Warranties: say yes, sometimes." ConsumerReports.org 70(1): 51.

DellaVigna, Stefano and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review* 96(3): 694 – 719.

DellaVigna, Stefano, Attila Lindner, Balazs Reizer, and Johannes Schmieder. "Reference-Dependent Job Search: Evidence from Hungary." Unpublished Working Paper.

Eliaz, Kfir and Ran Spiegler, 2011. "Consideration Sets and Competitive Marketing." *Review of Economic Studies* 78 (1): 235 – 262.

Ellison, Glenn. 2005. "A Model of Add-On Pricing", *The Quarterly Journal of Economics* 120(2): 585 – 637.

Ellison, Glenn and Sara Ellison. 2009. "Search, Obfuscation, and Price Elasticities on the Internet." *Econometrica* 77 (2): 427 – 452.

Ericson, Keith. 2014. "On the Interaction of Memory and Procrastination: Implications for Reminders." NBER Working Paper No. 20381.

Finkelstein, Amy. 2009. "E-ztax: Tax Salience and Tax Rates." *Quarterly Journal of Economics* 124(3): 969 – 1010.

Gabaix, Xavier. 2014. "A Sparsity-Based Model of Bounded Rationality." *Quarterly Journal of Economics*, forthcoming.

Gabaix, Xavier and David Laibson. 2006. "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets." *Quarterly Journal of Economics* 121 (2): 505 – 540.

Genesove, David and Christopher Mayer. 2001. "Loss Aversion and Seller Behavior: Evidence from the Housing Market." *Quarterly Journal of Economics* 116(4): 1233 – 1260.

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

Gennaioli, Nicola, Andrei Shleifer and Robert Vishny. 2012. "Neglected Risks, Financial Innovation and Financial Fragility." *Journal of Financial Economics* 104 (3): 452 – 468.

Gilboa, Itzhak and David Schmeidler. 1995. "Case-Based Decision Theory." *Quarterly Journal of Economics* 110(3): 605 – 639.

Gilchrist, Duncan, Michael Luca, and Deepak Malhotra. 2014. "When $3 + 1 > 4$: Gift Structure and Reciprocity in the Field." Working Paper, HBS.

Goldin, Jacob and Tatiana Homonoff. 2013. "Smoke Gets in Your Eyes: Cigarette Tax Salience and Regressivity." *American Economic Journal: Economic Policy* 5(1): 302 – 336.

Grubb, Michael. 2009. "Selling to Overconfident Consumers." *American Economic Review* 99(5): 1770 – 1807.

Grubb, Michael. 2014. "Consumer Inattention and Bill-Shock Regulation." *Review of Economic Studies* 82(1): 219 – 257.

Hanna, Rema, Sendhil Mullainathan and Joshua Schwartzstein. 2014. "Learning Through Noticing: Theory and Evidence from a Field Experiment" *Quarterly Journal of Economics* 129(3): 1311 – 1353.

Hastings, Justine and Jesse Shapiro. 2013. "Fungibility and Consumer Choice: Evidence from Commodity Price Shocks". *Quarterly Journal of Economics* 128(4): 1449 – 1498.

Hauser, John and Birger Wernerfelt. 1990. "An Evaluation Cost Model of Consideration Sets". *Journal of Consumer Research*, 16 (4): 393 – 408.

Hogarth, Robin, Hillel Einhorn. 1992. "Order Effects in Belief Updating: The Belief-Adjustment Model." *Cognitive Psychology* 24(1): 1 – 55.

Hsee, Christopher. 1998. "Less is better: When low-value options are valued more highly than high-value options." *Journal of Behavioral Decision Making* 11(2): 107 – 121.

Huysentruyt, Marieke and Daniel Read. 2010. "How do people value extended warranties? Evidence from two field surveys." *Journal of Risk and Uncertainty* 40(3): 197 – 218.

Jacoby, Larry, James Debner, and Janine Hay. 2001. "Proactive Interference, Accessibility Bias, and Process Dissociations: Valid Subjective Reports of Memory." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27(3): 686 – 700.

Kahneman, Daniel. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review* 93(5): 1449 – 1475.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux.

Kahneman, Daniel and Dale Miller. 1986. "Norm Theory: Comparing Reality to its Alternatives", *Psychological Review*, 93 (2): 136–153.

Kahneman, Daniel and Amos Tversky. 1973. "On the Psychology of Prediction." *Psychological Review* 80(4): 237 – 251.

Kahneman, Daniel and Amos Tversky. 1979. "Prospect Theory: an Analysis of Decision under Risk", *Econometrica*, 47 (2): 263– 292.

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman. 2014. "Getting to the Top of Mind: How Reminders Increase Saving," NBER Working Paper w16205.

Koszegi, Botond and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences", *Quarterly Journal of Economics*, 121 (4): 1133 – 1165.

Koszegi, Botond and Matthew Rabin. 2007. "Reference-Dependent Risk Attitudes", *American Economic Review*, 7 (4): 1047 – 1073.

Koszegi, Botond and Adam Szeidl. 2013. "A Model of Focusing in Economic Choice." *Quarterly Journal of Economics* 128 (1): 53 – 104.

Laibson, David. 2001. "A Cue-Theory of Consumption." *Quarterly Journal of Economics* 116(1): 81 – 119.

Malmendier, Ulrike and Stefan Nagel. 2011. "Depression Babies: Do Macroeconomic Experiences Affect Risk-Taking?" *Quarterly Journal of Economics* 126(1): 373 – 416.

Malmendier, Ulrike and Stefan Nagel. 2014. "Learning from Inflation Experiences." *Quarterly Journal of Economics*, conditionally accepted.

Mazumdar, Tridib, Raj S.P. and Indrajit Sinha. 2005. "Reference Price Research: Review and Propositions." *Journal of Marketing* 69(4): 84 – 102.

Mullainathan, Sendhil. 2002. "Memory-Based Model of Bounded Rationality." *Quarterly Journal of Economics* 117 (3): 735 – 774.

Mullainathan, Sendhil, Joshua Schwartzstein, and William Congdon. 2012. "A reduced-form approach to behavioral public finance." *Annual Reviews of Economics* 4 (17): 1 – 30.

Nedungadi, Prakash. 1990. "Recall and Consumer Consideration Sets: Influencing Choice without Altering Brand Evaluations." *Journal of Consumer Research* 17(3): 263 – 276.

Odean, Terrance. 1998. "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance* 53(5): 1775 – 1798.

Patrick, Kevin, Fred Raab, Marc Adams, Lindsay Dillon, Marian Zabinski, Cheryl Rock, William Griswold, and Gregory Norman. 2009. "Text Message-Based Intervention for Weight Loss: Randomized Controlled Trial." *Journal of Medical Internet Research*, 11.

Roberts, John and James Lattin. 1997. "Consideration: Review of Research and Prospects for Future insights", *Journal of Marketing Research*, 34 (3): 406 – 410.

Schwartzstein, Joshua. 2014. "Selective Attention and Learning", *Journal of the European Economic Association* 12(6): 1423 – 1452.

Sexton, Steven. 2014. "Automatic Bill Payment and Salience Effects: Evidence from Electricity Consumption." *Review of Economics and Statistics*, forthcoming.

Shocker, Allan, Moshe Ben-Akiva, Bruno Boccara and Prakash Nedungadi. 1991. "Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models, and Suggestions." *Marketing Letters* 2(3): 181 – 197.

Simonsohn, Uri and George Loewenstein, 2006. "Mistake #37: The Effect of Previously Encountered Prices on Current Housing Demand." *Economic Journal* (116)508: 175 – 199.

Simonson, Itamar, and Amos Tversky. 1992. "Choice in Context: Tradeoff Contrast and Extremeness Aversion", *Journal of Marketing Research*, 29 (3): 281 – 295.

Stango, Victor and Jonathan Zinman. 2014. "Limited and Varying Consumer Attention: Evidence from Shocks to the Salience of Bank Overdraft Fees." *Review of Financial Studies* 27(4): 990 – 1030.

Sussman, Abigail and Adam Alter. 2012. "The Exception Is the Rule: Underestimating and Overspending on Exceptional Expenses" *Journal of Consumer Research* 39(4): 800 – 814.

Taubinsky, Dmitry. 2014. "From Intentions to Actions: A Model and Experimental Evidence of Inattentive Choice." Mimeo Harvard University.

Taylor, Shelley and Suzanne Thompson. 1982. "Stalking the ellusive vividness effect." *Psychological Review*, 89 (2): 155 – 181.

Tversky, Amos and Daniel Kahneman. 1973. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5(2): 207 – 232.

Tversky, Amos and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185(4157): 1124 – 1131.

Tversky, Amos, Paul Slovic and Daniel Kahneman. 1990. "The Causes of Preference Reversal." *American Economic Review* 80(1): 204 – 217.

Vervloet, Marcia, Annemiek Linn, Julia van Weert, Dinny de Bakker, Marcel Bouvy, and Liset van Dijk. 20120. "The Effectiveness of Interventions Using Electronic Reminders to Improve Adherence to Chronic Medication: A systematic Review of the Literature." *Journal of the Americal Medical Informatics Association* 19: 696 – 704.

A Proofs

Lemma 1. Because $q_1 - p_1 = q_2 - p_2$, a FAST consumer (with $\delta < 1$) chooses the high quality apartment 1 whenever its quality is salient, and chooses apartment 2 otherwise. (Note that equal salience for quality and price is excluded by the assumption that 1 has strictly higher quality and higher price than 2.) To see this note that $q_1 - q_2 > \delta(p_1 - p_2)$, so if 1 is quality salient it is chosen (even if 2 is also quality salient), and that $\delta(q_1 - q_2) > p_1 - p_2$, if 1 is price salient than it is not chosen (even if 2 is also price salient). Intuitively, salience distortions are more important for the option with higher attribute levels.

For a consumer with normal prices $p_k^n = p_k + \Delta$, the reference price is $\overline{p} = \frac{p_1+p_2+\Delta}{2}$. Thus, for $\Delta \in [-(p_1 - p_2), (p_1 - p_2)]$ the reference price is in the range $[p_2, p_1]$. For any such Δ, apartment 1 is weakly more expensive than the reference and apartment 2 is weakly cheaper than the reference. Thus, the quality of apartment 1 is salient when

$$\frac{q_1}{(q_1 + q_2)/2} > \frac{p_1}{(p_1 + p_2 + \Delta)/2}$$

namely when $\Delta > \Delta^* = p_1 \left(\frac{q_2}{q_1} - \frac{p_2}{p_1} \right)$ where $\Delta^* > 0$ by assumption.

Consider now the case where $\Delta < -(p_1 - p_2)$. Then price is still salient for apartment 1, so apartment 2 is chosen. Consider finally the case where $\Delta > p_1 - p_2$. In this case, apartment 1 is strictly cheaper than the reference price, while being of higher than average quality. Thus, its quality is salient if and only if

$$\frac{q_1}{(q_1 + q_2)/2} > \frac{(p_1 + p_2 + \Delta)/2}{p_1}$$

which reads $\Delta < \Delta^{**} = \frac{3\frac{q_1}{q_2} - 1}{\frac{q_1}{q_2} + 1} p_1 - p_2$. For price shifts larger than Δ^{**}, the recalled prices are so much higher than the observed prices that the difference between the choice set prices and the evoked set prices is very salient. As a consequence, the consumer undervalues apartment 1 relative to apartment 2 and chooses the latter. ■

Lemma 2. Consider consumer *Pit*. The normal price for apartment $k = 1, 2$ is p_k since it has $\pi_k(p_k) = e_k(p_k) = 1$ so that $a(1, 1) = 1$ and p_k is recalled for any $a^* \in (0, 1]$. Moreover, no other price is recalled since for any $p \neq p_k$ we have $\pi_k(p') = e_k(p') = 0$ and $a(0,0) = 0 < a$.

Consider now consumer SF. Exposure to prices for apartment k is $e_k(p) = 1$ for $p = p_k + \Delta$ and zero otherwise. Frequency of prices for apartment k is $\pi_k(p) = 1$ for $p = p_k$ and zero otherwise. Thus, for apartment k, price $p_k + \Delta$ has availability $a(0,1)$ and price p_k has availability $a(1,0)$. All other prices have availability zero and are not recalled. As a consequence, for $a^* \leq a(0,1)$ the consumer recalls past prices $p_k + \Delta$ for both apartments $k = 1,2$. In this case, normal prices satisfy

$$p_k^{n,SF} = \begin{cases} p_k + \frac{a(0,1)}{a(0,1)+a(1,0)}\Delta & \text{if } a^* \leq a(1,0) \\ p_k + \Delta & \text{if } a^* > a(1,0) \end{cases}$$

and so are strictly higher than p_k for positive Δ.

If instead $a^* > a(0,1)$ then the consumer does not bring to mind past prices. If $a^* \leq a(1,0)$ then he (only) recalls current prices and sets $p_k^{n,SF} = p_k$. If $a^* > a(1,0)$ no price comes to mind, and the consumer takes the visible price as normal. In this case, we again have $p_k^{n,SF} = p_k$. ■

Proposition 1. Consider first an FBOR consumer. Because $u^{FBOR}(q_i, p_{ui}) = q_i - p_{ui}$ and $u^{FBOR}(q_i, p_{ui}, p_{m_i}) = q_i - p_{ui} - \mu p_{mi}$ for $i = 1,2$, it follows that $\Delta u_{1,2}^{FBOR}(q_i, p_{ui}, p_{m_i}) = \Delta u_{1,2}^{FBOR}(q_i, p_{ui}) + \mu(p_{m1} - p_{m,2})$.

Consider now a FAST consumer. Without loss of generality, assume that $q_1 > q_2$, $p_{u1} > p_{u2}$ and $p_{m1} > p_{m2}$, as well as $q_1/q_2 > p_{u1}/p_{u2}$. Then $u^{FAST}(q_i, p_{ui}) = \frac{2}{1+\delta}(q_i - \delta p_{ui})$ for $i = 1,2$. Consider two cases:

- If $p_{m1}/p_{m2} > q_1/q_2$, then $u^{FAST}(q_i, p_{ui}, p_{m,i}) = \frac{2+\mu}{\delta+\delta^2+\mu}(\delta q_i - \delta^2 p_{ui} - \mu p_{m,i})$ for $i = 1,2$. As a consequence,

$$\Delta u_{1,2}^{FAST}(q_i, p_{ui}, p_{m_i}) = \frac{2+\mu}{2} \cdot \frac{1+\delta}{\delta+\delta^2+\mu} \Delta u_{1,2}^{FAST}(q_i, p_{ui}) + \frac{2+\mu}{\delta+\delta^2+\mu} \cdot \mu(p_{m1} - p_{m,2})$$

Under the assumption that, prior to information disclosure, FAST prefers the high quality good, namely $\Delta u_{1,2}^{FAST}(q_i, p_{ui}) > 0$, we find that

$$\Delta u_{1,2}^{FAST}(q_i, p_{ui}, p_{m_i}) - \Delta u_{1,2}^{FAST}(q_i, p_{ui}) < -\frac{2+\mu}{\delta+\delta^2+\mu} \cdot \mu(p_{m1} - p_{m,2}) < -\mu(p_{m1} - p_{m,2})$$

so FAST's valuation drops by more than FBOR's valuation upon being informed about a salient cost.

- If $p_{m1}/p_{m2} < p_{u,1}/p_{u,2}$, then $u^{FAST}(q_i, p_{ui}, p_{m,i}) = \frac{2+\mu}{1+\delta+\delta^2\mu}(q_i - \delta p_{ui} - \delta^2 \mu p_{m,i})$ for $i = 1, 2$. As a consequence,

$$\Delta u_{1,2}^{FAST}(q_i, p_{ui}, p_{m_i}) = \frac{2+\mu}{2} \cdot \frac{1+\delta}{1+\delta+\delta^2\mu} \Delta u_{1,2}^{FAST}(q_i, p_{ui}) + \frac{2+\mu}{1+\delta+\delta^2\mu} \cdot \delta^2 \mu(p_{m1} - p_{m,2})$$

Under the assumption that, prior to information disclosure, FAST prefers the high quality good, namely $\Delta u_{1,2}^{FAST}(q_i, p_{ui}) > 0$, we find that

$$\Delta u_{1,2}^{FAST}(q_i, p_{ui}, p_{m_i}) - \Delta u_{1,2}^{FAST}(q_i, p_{ui}) > -\frac{2+\mu}{1+\delta+\delta^2\mu} \cdot \delta^2 \mu(p_{m1} - p_{m,2}) > -\mu(p_{m1} - p_{m,2})$$

so FAST's valuation drops by more than FBOR's valuation upon being informed about a salient cost.

■

Proposition 2. TVs are characterized by three attributes, q, p, f, where q is (visible) quality of a working TV, p is (visible) price and f is (shrouded) quality of TV in the contingency of a breakdown. For a rational consumer (with full recall and no salience distortions) the utility of TV (q, p, f) is $\theta_1 q + (1 - \theta_1) f - p$. Below, we assume $f = 0$.

Let $C = C(n) = \{(q, q, p'), (q, 0, p), (0, 0, 0)\}$. The reference attributes are $\overline{q} = q\frac{2}{3}$, $\overline{f} = q\frac{1}{3}$ and $\overline{p} = \frac{p+p'}{3}$. It is easy to see that for p' not much larger than p, the perceived utilities of the two options (TV with and without insurance) are:

$$\begin{aligned} u^S(q, q, p') &= \frac{1}{\theta_1 \delta^2 + (1-\theta_1) + \delta} \left[(\theta_1 \delta^2 + 1 - \theta_1) \cdot q - \delta \cdot p' \right] \\ u^S(q, 0, p) &= \frac{1}{\theta_1 \delta + (1-\theta_1) + \delta^2} \left[\theta_1 \delta \cdot q - \delta^2 \cdot p \right] \end{aligned}$$

namely, the breakdown contingency is the most salient attribute for both options, but price is more salient for the more expensive TV with insurance than for the baseline option without insurance. The outside option gives utility zero.

It is easy to see that for $\delta \to 0$, we get $u^S(q, q, p') = q$ which is larger than $u^S(q, 0, p) = u^S(0, 0, 0) = 0$. This shows that, in this limit, the consumer buys the TV with insurance for arbitrary large prices. We divide the analysis in two stages: we first show that, for given baseline price $p > 0$, there exists δ^* such that for $\delta < \delta^*$ the consumer's willingness to pay for the insured TV satisfies $WTP > p + (1 - \theta_1)q$, i.e. the consumer is willing to overpay for insurance *relative* to the uninsured option. [..]

We then address directly the incentives to unshroud the breakdown risk, by asking the question: relative to the maximum price the seller can charge when f is shrouded, i.e. $p = q$, can the seller increase his profits by unshrouding the breakdown risk and offering insurance? We show that, even if the cost of providing insurance equals the actuarially fair price $(1-\theta_1)q$, it is profitable for the seller to unshroud for δ small enough. [...] ■

Proposition 3. As shows in the text, upon unshrouding the valuation of FAST in the "tax unshrouding perspective" drops from $q - p$ to $q - p(1 + \tau)$. In contrast, the valuation of FAST in the "price surprise" perspective drops from $q - p$ to $\frac{2}{1+\delta}[\delta q - p(1 + \tau)]$. The excess reduction in valuation is

$$\frac{2}{1+\delta}[\delta q - p(1 + \tau)] - [q - p(1 + \tau)] = -\frac{1-\delta}{1+\delta}[q + p(1 + \tau)]$$

■

B Frequency based recall

Consider the case in which recall is only driven by likelihood. A notable example is a long run situation in which the consumer has sampled prices for many periods, so that his past exposure to a price coincides with its frequency.[35] Formally, in this case $e_{p_k}(p) = \pi_{p_k}(p)$ for

[35] Recall is also driven by frequency in other interesting cases, such as when consumers have no exposure but have been given a description of the price distribution prior to choice. Also in this case all price realisations receive the same exposure.

any price p of good k. The availability of p is then an increasing function $a\left(\pi_{p_k}(p), \pi_{p_k}(p)\right)$ of its frequency $\pi_{p_k}(p)$.

To characterize recall in this situation, note first that, because the price of good 2 is assumed to be deterministic, availability of p_2 is maximal, $a\left(1,1\right) = 1$. The normal price of good 2 is its actual price, $p_2^n = p_2$, for any availability threshold $a^* \leq 1$.

It is also easy to characterize the normal price of good 1. Because the availability of price p increases in its frequency, the availability threshold a^* pins down a likelihood threshold π_1^*. All price realizations p_1 that occur with frequency greater than or equal to π_1^* are then recalled; those occurring with frequency is below π_1^* are neglected. In what follows, we restrict the analysis to the interesting case in which a^* is sufficiently low relative to the mode of $\pi_{p_1}(p)$ that some price realisations are recalled, so the set $X_{p_1}(a^*) = \{p \,|\pi_{p_1}(p) \geq \pi_1^*\}$ is non empty. The normal price for good 1 is then given by:

$$p_1^n = \mathbb{E}_{\pi_{p_1}} \left\{p \,|\pi_{p_1}(p) \geq \pi_1^* \right\}. \tag{14}$$

When recall is driven by frequency, the price that the consumer considers normal can coincide with the average price, but only when the distribution $\pi_{p_1}(p)$ is symmetric around the mean. If for instance the price distribution $\pi_{p_1}(p)$ is skewed, the normal price is distorted towards the modal price, capturing the intuition that consumers neglect tail price realisations.

Equation (14) has interesting implications for how the interaction between memory and attention works under different attribute distributions. We proceed under the conditions of Lemma 1, namely that good 1 is always more expensive than the average good, namely $\underline{p} \geq \left(p_1^n + 2p_2^n\right)/3$.

Proposition 4 *Suppose that $\pi_{p_1}(\cdot)$ is unimodal, with mode p_1^m, and that $\pi_{p_1}(\cdot)$ is sufficiently concentrated around p_1^m that $\pi_{p_1}(p_2\frac{q_1}{q_2}) < \pi_1^*$. We then have two cases:*

i) If p_1^m is low, $p_1^m < p_2\frac{q_1}{q_2}$, then quality is salient at the likely prices around p_1^m. The normal price satisfies $p_1^n < p_2\frac{q_1}{q_2}$ and price becomes salient only for low probability, right tail, prices $p_1 > \widehat{p} > p_1^m$.

ii) If p_1^m is high, $p_1^m > p_2\frac{q_1}{q_2}$, then price is salient at the likely prices around p_1^m. The normal price satisfies $p_1^n > p_2\frac{q_1}{q_2}$ and quality becomes salient only for low probability, left tail, *prices* $p_1 < \widehat{p} < p_1^m$.

Proof. If π_{p_1} is unimodal, then for $a^* > 0$ the set of recalled prices is a closed interval $X_{p_1}(a^*) = [p_{\min}(a^*), p_{\max}(a^*)]$. In particular, the normal price $p_1^n = \mathbb{E}_{p_1}[p_1|p_1 \in X_{p_1}(a^*)]$ satisfies $p_1^n \in X_{p_1}(a^*)$. Moreover, if π_{p_1} is sufficiently concentrated around its mode p_1^m that $\pi_{p_1}\left(p_2 \cdot \frac{q_1}{q_2}\right) < \pi^*$ (where π^* is the threshold probability satisfying $a(\pi^*, \pi^*) = a^*$), then $p_2 \cdot \frac{q_1}{q_2}$ lies outside $X_{p_1}(a^*)$. Consider two cases:

Case i), $p_1^m < p_2 \cdot \frac{q_1}{q_2}$. Then $p_2 \cdot \frac{q_1}{q_2} > p_{\max}(a^*) > p_1^n$. Quality is salient for p_1 in a neighbourhood of p_1^n. If π_{p_1} is sufficiently concentrated around its mode, this neighborhood includes the modal price.

Case ii), $p_1^m > p_2 \cdot \frac{q_1}{q_2}$. Then $p_2 \cdot \frac{q_1}{q_2} < p_{\max}(a^*) < p_1^n$. Price is salient for p_1 in a neighbourhood of p_1^n. If π_{p_1} is sufficiently concentrated around its mode, this neighborhood includes the modal price. ■

In stable unimodal settings, FASTs behave like FBORs most of the time, in the sense that they display a stable quality-price tradeoff for the frequent prices around the mode. This quality-price tradeoff is shaped by salience, not just by underlying preferences, but this influence would not be detectable by an external observer. To see this, consider case i), in which the modal price of the high quality good is not so high (relative to the alternative). Here quality is salient at normal prices, and FAST behaves most of the times like a rational consumer with a high taste for quality. On the other hand, in case ii) price is salient at the high normal price, and FAST behaves most of the times like a rational consumer with a low taste for quality.

For stable and unimodal distributions, FAST departs from FBOR only in the low probability event that an extreme price occurs and surprises him. In this case, the quality-price tradeoff of FAST changes, while it would remain stable for FBOR. When quality is salient at normal prices (case i), the realization of an unlikely high price causes a shift to price salience, drastically increasing the price sensitivity of FAST. When price is salient at normal prices (case ii), the realization of an unlikely low price causes a shift to quality salience, drastically reducing the price sensitivity of FAST.

In sum, for unimodal distributions, FAST displays stable preferences most of the time but overreacts to price changes that are extreme and therefore, given uni-modality, rather unlikely. In contrast, if the price distribution is bimodal, instability of preferences can be much more common. This intuition is captured by the following result:

Corollary 1 *Suppose that $\pi_{p_1}(\cdot)$ has two local maxima (or modes) $p^m_{1,L}$ and $p^m_{1,H}$, with $p^m_{1,L} < p_2\frac{q_1}{q_2} < p^m_{1,H}$. If $\pi_{p_1}(\cdot)$ is sufficiently concentrated around its modes that $p^n_1 \in \left(p^m_{1,L}, p^m_{1,H}\right)$, then quality is salient at the likely prices around $p^m_{1,L}$ while price is salient at the likely prices around $p^m_{1,H}$.*

Proof. It is sufficient to show that $\hat{p} = p^n_1 \frac{\frac{1}{2}+\frac{p_2}{p^n_1}}{\frac{1}{2}+\frac{q_2}{q_1}}$ satisfies $\hat{p} \in (p^m_{1L}, p^m_{1H})$, when both p^n_1 and $p_2\frac{q_1}{q_2}$ lie inside the same interval. First, note that $\hat{p} > p^m_{1L}$ can be rewritten $\frac{1}{2}\left(p^n_1 - p^m_{1L}\right) + \left(p_2 - \frac{q_2}{q_1}p^m_{1L}\right) > 0$. The inequality holds because both terms are positive, given the assumption that p^n_1 and $p_2\frac{q_1}{q_2}$ are larger than p^m_{1L}. Similarly, $\hat{p} < p^m_{1H}$ can be rewritten $\frac{1}{2}\left(p^n_1 - p^m_{1H}\right) + \left(p_2 - \frac{q_2}{q_1}p^m_{1H}\right) < 0$, which holds because both terms are negative, given the assumption that p^n_1 and $p_2\frac{q_1}{q_2}$ are smaller than p^m_{1H}. ■

When the price distribution is bi-modal and the expected price is in between two modes, price is salient for the high price mode $p^m_{1,H}$ while quality is salient for the low price mode $p^m_{1,L}$. Attention switches frequently. Realizations around the high price mode look surprising compared to the low price mode, and thus attract the consumer's attention to the higher price of good 1. By contrast, realizations around the low price mode look surprising compared to the high price mode, attracting the consumer's attention to the higher quality of good 1. The preferences of FAST are maximally unstable, in the sense that he is similarly likely to be price sensitive and price insensitive.

We now illustrate the implications of this model in the context of sales. In BGS (2013) we showed that salient thinking can rationalize the phenomenon of "misleading sales". This term denotes strategic price reductions that surprise consumers, render quality salient and boosting consumers' willingness to pay for the good. In BGS (2013) we studied sales when consumers hold rational expectations of price, namely $p^n_1 = \mathbb{E}(p_1)$. In the current, more general, setting, normal prices may depart from rationally expected prices.

A consumer chooses between good (q_1, p_1) and the outside option $(q_2, p_2) = (0, 0)$ of not buying anything. The choice is repeated over time, and the consumer has been randomly exposed to different realizations of an underlying price distribution so that exposure and frequency coincide. In this setting, we compare two pricing strategies by firm 1: i) a deterministic price policy specifying a fixed price p_1, and ii) a sales policy setting a regular price p_h with probability $1 - \pi_l$ and a lower sale price $p_l < p_h$ with probability π_l. The appendix then proves the following result.

Proposition 5 *The seller's revenue depends on its price policy:*

i) The maximum deterministic price for which revenue is positive is $p_1^{\max} = q_1$. *At this price, the firm always sells the good and obtains the per period revenue* $r^{det} = q_1$.

ii) A stochastic "sales" policy setting an inflated regular price $p_h \in (q_1/\delta, 7q_1/\delta)$ *and a sale price* $p_l = \frac{q_1}{\delta}$ *yields the per period expected revenue* r^{sales} *as a function of sales probability* π_l, *where:*

$$r^{sales} = \begin{cases} \pi_l \frac{q_1}{\delta} & \text{for} \quad \pi_l < 1 - \pi_1^* \\ \\ 0 & \text{for} \quad \pi_l \geq 1 - \pi_1^* \end{cases},$$

and $\pi_1^* < 1/2$ *is the recall threshold. The firm sells the good only at the sale price* p_l *and provided the consumer recalls the regular price* p_h.

Proof. When price p_1 is deterministic, we have $C = C(n) = \{(q_1, p_1), (0, 0)\}$. The reference quality is $\overline{q} = q_1/2$ and the reference price is $\overline{p} = p_1/2$. For any price p_1, quality and price are equally salient, $\sigma(q_1, \overline{q}) = \sigma(1, 1/2) = \sigma(p_1, \overline{p})$. The consumer's valuation for good 1 in this context is then $u^S(q_1, p_1) = q_1 - p_1$, while his valuation for the outside option is zero. The consumer's willingness to pay is then $WTP = \max\{p_1 : q_1 - p_1 > 0\} = q_1$.

Suppose now price p_1 is stochastic, with $p_1 = p_l$ with probability π_l and $p_1 = p_h$ with probability $1 - \pi_l$. Given the assumption that exposure to prices coincides with their relative frequency, $e_{p_1}(p) = \pi_{p_1}(p)$ for all p, it follows that a price realization p is recalled if and only if its frequency is $\pi_{p_1}(p) \geq \pi_1^*$, where $a(\pi_1^*, \pi_1^*) = a^*$.

Given the assumption that $\pi_1^* < 1/2$, the normal price p_1^n recalled by consumers under this price policy is:

$$p_1^n = \begin{cases} p_h & \text{if} \quad \pi_l < \pi_1^* \\ \mathbb{E}(p_1) & \text{if} \quad \pi_l \in [\pi_1^*, 1-\pi_1^*] \\ p_l = \frac{q_1}{\delta} & \text{if} \quad \pi_l > 1-\pi_1^* \end{cases} .$$

To proceed, take $p_l = q_1 \cdot \frac{1}{\delta}$ and $p_h \in \left(q_1 \cdot \frac{1}{\delta}, q_1 \cdot \frac{7}{\delta}\right)$. First, note that the consumer's maximum WTP for good 1 is p_l. This occurs when at the price p_l the good's quality is salient. Because $p_h > p_l$, the consumer never buys the good at the regular price.

We now examine when the consumer buys at the sales price p_l, namely when the choice set is $C = \{(q_1, p_l), (0,0)\}$. The evoked set is $C(n) = \{(q_1, p_1^n), (0,0)\}$ (where p_1^n is given above), and the reference good is $\overline{q} = q_1/2$, $\overline{p} = (p_l + p_1^n)/4$. For good 1 the salience of quality is $\sigma(q_1, \overline{q}) = \sigma(1, 1/2)$ and that of price is $\sigma(p_l, \overline{p}) = \sigma(1, \overline{p}/p_l)$. There are three cases:

- $p_1^n = p_h$. In this case, $\overline{p} = (p_h + p_l)/4$. Then, if $p_l > \overline{p}$, that is if $p_l \geq p_h/3$, we have $\sigma(q_1, \overline{q}) \geq \sigma(p_l, \overline{p})$ iff $\frac{q_1}{\overline{q}} > \frac{p_l}{\overline{p}}$ which always holds since $p_h > p_l$. If instead $p_l < \overline{p}$, that is if $p_l < p_h/3$, then good 1's quality is salient iff $p_l > p_h/7$. Thus, when $p_1^n = p_h$, quality is salient provided $p_h \in (p_l, 7p_l)$.

- $p_1^n = \mathbb{E}(p_1)$. In this case, $\overline{p} = (p_h + p_l)/4$. Then, if $p_l > \overline{p}$, that is if $p_l > \frac{1-\pi_l}{3-\pi_l} p_h$, we have $\sigma(q_1, \overline{q}) \geq \sigma(p_l, \overline{p})$ iff $\frac{q_1}{\overline{q}} > \frac{p_l}{\overline{p}}$, which always holds. If instead $p_l < \overline{p}$, that is if $p_l < \frac{1-\pi_l}{3-\pi_l} p_h$, then good 1's quality is salient iff $p_l > \frac{1-\pi_l}{7-\pi_l} p_h$. Thus, when $p_1^n = \mathbb{E}(p_1)$, quality is salient provided $p_h \in \left(p_l, \frac{7-\pi_l}{1-\pi_l} p_l\right)$. Note that $\frac{7-\pi_l}{1-\pi_l} > 7$, so this interval is larger than the one in the previous case.

- $p_1^n = p_l$. In this case, $p_l > \overline{p} = p_l/2$. As a consequence, good 1's quality and price are equally salient, $\sigma(q_1, \overline{q}) = \sigma(p_l, \overline{p}) = \sigma(1, 1/2)$, so quality is never strictly salient.

In short, for $p_l = q_1/\delta$ and $p_h \in (p_l, 7p_l)$, quality is salient at price p_l if and only if the normal price is $p_1^n = p_h$ or $p_1^n = \mathbb{E}(p_1)$. This bounds the sales frequency to $\pi_l \leq 1 - \pi^*$, so the maximum revenues are equal to $\frac{1-\pi^*}{\delta} q_1$. ■

Consider first the deterministic price policy, case i). Here the normal price of good 1 is p_1 itself. The choice context is $C(n) \cup C = \{(q_1, p_1), (q_1, p_1), (0,0), (0,0)\}$. In this context, the reference attribute levels are $\overline{q} = q_1/2$ and $\overline{p} = p_1/2$. As a consequence, the price and quality of good 1 are equally salient, for $\sigma(q_1, q_1/2) = \sigma(p_1, p_1/2) = \sigma(1, 1/2)$, FAST values good 1 rationally and he is willing to pay at most $p_1 = q_1$ for it.

By implementing a stochastic price, by contrast, the firm can render quality salient during the sales event. This allows the firm to extract a sale price equal to the consumer's perceived valuation under salient quality, namely $p_l = q_1/\delta$. For this to happen, though, the sales event must be a surprising price reduction relative to the normal price. With our recall process, the normal price p_1^n recalled by consumers under this price policy is:

$$p_1^n = \begin{cases} p_h & \text{if} \quad \pi_l < \pi_1^* \\ \mathbb{E}(p_1) & \text{if} \quad \pi_l \in [\pi_1^*, 1 - \pi_1^*] \\ p_l = \frac{q_1}{\delta} & \text{if} \quad \pi_l > 1 - \pi_1^* \end{cases} .$$

When the probability of sales is low, the normal price recalled by the consumer is the high regular price p_h. This is similar to case ii) of Proposition 4: as the price drops to p_l the consumer's attention is drawn to quality, and FAST buys the good. When both prices are reasonably likely, the consumer recalls both and the normal price is the rationally expected price $\mathbb{E}(p_1)$, as in BGS (2013). This captures the bimodal case of Corollary 2. In this case, price is salient at the regular price but quality is salient at the sales price, which induces FAST to buy the good. Finally, when sales are very frequent, the normal price is the sale price. Here, a low price is normal.[36]

The changing normal price has key implications for the effectiveness of the sales policy. When sales are infrequent enough, $\pi_l < 1 - \pi_1^*$, sales render quality salient and entice the consumer to pay the inflated sales price $p_l = q_1/\delta$. If however sales become too frequent, $\pi_l > 1 - \pi_1^*$, the normal price becomes the sale price itself. As a consequence, the sale price $p_l = q_1/\delta$ is no longer surprising, quality is no longer salient, and FAST chooses not to buy the good. Here sales backfire, yielding zero revenue.

[36] Conversely, as recall becomes more limited and π_1^* increases towards 1/2, the consumer no longer recalls the less frequent price. In this way, as limited recall gets more severe, the consumer's notion of normal price gets more distorted. This in turn sets the stage for a surprising price realization.

Relative to a deterministic price policy $p_1 = q_1$ that extracts the consumer's rational willingness to pay, a "misleading sales" policy entails the following tradeoff. On the one hand, during sales events the misleading sale boosts consumer willingness to pay up to $p_l = q_1/\delta$. On the other hand, sales work only if they are sufficiently unlikely so that normal prices are high and consumers are, in fact, surprised by low prices. Interestingly, our analysis indicates that at the profit maximizing sales strategy $\pi_l = 1 - \pi_1$, the consumer expects the sales event to happen, but the comparatively low sales price renders quality salient nonetheless.

Evidence from the marketing literature provides some support for the predictions of our model: regular and frequent sales drastically lower consumers' price "standard", against which they evaluate other prices (Mazumdar, Raj and Sinha, 2005), and reduces demand at regular prices (REFS). In the above analysis we highlighted the frequency of sales as a determinant of normal prices. In practice, customers self selecting into sales events would be disproportionately exposed to the sales price, and thus view it as normal, regardless of the actual frequency of sales. This consideration suggests that exposure can be an important driver of recall. In the next Section, we explore more closely the role of past exposure on choice behavior.

C Dinnerware Sets and Broken Dishes

The distinction between FAST and FBOR can shed light on Christopher Hsee's famous dinnerware set experiment. In this experiment, subjects stated higher willingness to pay for a dinnerware set consisting of 24 good dishes than for an identical dinnerware set consisting of 31 dishes plus a few broken ones. This result violates the basic economic principle that "more is better" (and free disposal). Consider now this experiment in the context of our model.

Suppose that subjects are asked to evaluate a dinnerware set consisting of $N > 1$ good dishes. Subjects must state their willingness to pay Np, where for convenience price p is in per-dish terms. As per Equation (1), the dinnerware set delivers utility:

$$Nq - Np,$$

where q is per-dish utility value.

When assessing the good (Nq, Np), subjects are choosing from the choice set $C(G) = \{(Nq, Np), (0, 0)\}$, where $(0, 0)$ is the option of not buying. A rational subject would state a willingness to pay $p = q$. A salient thinker departs from the reational benchmark because he recalls the normal version of his choice set $C(G, n) = \{(Nq, Np^n), (0, 0)\}$, where p^n is the normal price the subject is accustome to pay for dishes.

In the context $C(G) \cup C(G, n)$ the average good is $(Nq/2, N(p + p^n)/4)$. It is easy to see that this implies:

Lemma 3 *There are two cases:*

a) If the consumer is FBOR, $\delta = 1$, willingness to pay is equal to quality, $p = q$.

b) If the consumer is FAST, $\delta < 1$, then provided $p^n \in (\delta q, q]$ willingness to pay is equal to the normal recalled price, namely $p = p^n$.

For a rational consumer (there is no forgetting in this stage), willingness to pay is independent of the context created by normal price. As in BGS (2013), salience causes willingness to pay is anchored to the normal price. In fact, when $p = p^n$ price and quality are equally salient, and people are willing to pay up to q. The reason subjects are not willing to pay more than p^n is because at $p > p^n$ price becomes salient. As a result, valuation per dish drops to δq which is below p^n by the condition of Lemma 2.

Consider now a subject who is asked to state his willingness to pay for a dinnerware set consisting of $N + B$ dishes, $B > 0$ of which are broken. This dinnerware set contains the same amount N of good dishes. However, the number $B > 0$ of broken dishes brings another dimension into consideration. Formally, the consumer's utility from the dinnerware set depends on this new dimension as follows:

$$(N + B)q - Bq - (N + B)p.$$

The consumer effectively buys a dinenrware set of dimension $(N + B)$ but he must discount the B broken dishes from total value. In utility terms, this dinnerware set is a triplet

$((N+B)q, Bq, (N+B)p)$ capturing the total size $N+B$ of the dinnerware set, its broken dishes B, and the total price (again in per dish term).

The choice set is now equal to $C(G) = \{((N+B)q, Bq, (N+B)p), (0,0,0)\}$. As a result, a rational consumer states his willingness to pay to be equal to $p = qN/(N+B)$. Intuitively, the rational consumer is willing to pay q only for the good dishes. As a result, he is willing to pay the exact same price for a dinnerware set consisting of N good dishes and for a dinnerware set consisting of $N+B$ dishes B of which are broken.

Consider now the case of a salient thinker. When presented the dinnerware set $((N+B)q, Bq, (N+B)p)$, subjects recall a normal dinnerware set of the same size. Critically, now subjects do not need to fit only a normal price per dish p^n, but also a normal realization of the new (dis-) utility dimension, namely the number of broken dishes B^n. Of course, now the normal price per dish is the normal price of a dinnerware set with B^n broken dishes, which in the market it is in turn equal to $p^n N/(N+B^n)$, namely the normal price of good dishes p^n times the fraction of non broken dishes in the dinnerware set. In this new experiment, then, the normal version of the choice set is equal to $C(G,n) = \left\{((N+B)q, B^n q, \left(\frac{N+B}{N+B^n}\right) p^n N), (0,0,0)\right\}$.

It is easy to see that in this new choice context the result below obtains.

Proposition 6 *Consider the two following contrasting cases:*

a) The consumer recalls a dinnerware set with the same quality as the one at hand, consisting of exactly B broken dishes, namely $B^n = B$. In this case, FBOR states a willingness to pay equal to $p = qN$, while FAST states a willingness to pay $p = p^n N/(N+B)$. Both FBOR and FAST value the current dinnerware set the same as one with N good dishes and zero broken ones.

b) The consumer recalls a dinnerware set with no broken dishes, namely $B^n = 0$. In this case, FBOR continues to state a willingness to pay equal to $p = qN$. On the other hand, there exists a threshold $\underline{\delta}$ such that, for $\delta < \underline{\delta}$, FAST states a willingness to pay $p < p^n N/(N+B)$. That is, now FAST (and only FAST) can value the current dinnerware set less than one consisting of N good dishes and zero broken ones.

Our model yields the Hsee dinnerware set experiment provided two conditions holds. First, the normal dinnerware sets we recall has no broken dishes $B^n = 0$. Second, provided the consumer is FAST and salience distortions are strong enough, namely $\delta < \underline{\delta}$.

It is clear that the normal dinnerware sets we buy have no broken dishes, so the empirically plausible case is b), in which $B^n = 0$. Here, the logic of the result goes as follows. When seeing the dinnerware set with broken dishes, FAST values it by contrasting it with a normal dinnerware set, which has no broken dishes. In this comparison, the dimension of broken dishes is very salient. As the subject's attention is directed toward it, he overweights broken dishes in his willingness to pay, downard adjusting his wtp too much. Clearly, this result is not just reliant on recalling $B^n = 0$, but also on attention distortions being sufficiently strong. If salience causes no attention distortions, namely $\delta \geq \underline{\delta}$, the consumer does not pay disproprotionate attention to broken dishes. As a result, his stated price is rational.

The empirically implausible case a), on the other hand, underscores the critical role of recall. If the normal dinnerware set recalled by the consumer has the same amount of broken dishes as the dinnerware set at hand, namely $B^n = B$, broken dishes are not more salient than good dishes. As a result, even if salience distortions are in principle present (i.e., δ is small), attention is properly balanced among different attribute and wpt is rational.

This result underscore the interaction between memory and attention in generating surprises and overreaction. Seeing broken dishes is surprising in the very specific sense that, since broken dishes are very a-normal, they immediately attract the consumer's attention. This causes the consumer's overreaction to the new dimension of broken dishes that dramatically reduces his valuation below the rational counterpart.