Comparisons and Choice

Tom Cunningham[*]

17th November 2011

Abstract

Decisions made in laboratory settings have often been found to be sensitive to details of how the choice sets are structured and ordered. I provide a framework for describing these effects, in which preferences can vary depending on the history of choice sets, but which restricts them to be consistent when decisions are made simultaneously. Within this framework I introduce a simple comparison effect, in which being exposed to a larger value along some dimension makes a person less sensitive to marginal differences along that dimension (expressed in terms of the marginal rate of substitution). This effect predicts the existence of many otherwise disparate empirical findings (contrast and anchoring effects, response range effects, joint vs separate reversals, decoy, scope neglect and common difference effects). I also discuss sufficient conditions under which this behaviour could reflect Bayesian inference from the choice set. In the second half of the paper I apply the model to equilibrium in an imperfectly competitive market. Under unit demand the model predicts, unlike the rational benchmark, that markups and dispersion are increasing in cost, two facts that have been commonly observed in the empirical IO literature. Finally I introduce a novel dataset of cost and price for 3,500 goods from a chain drugstore which shows, as predicted, a very tight connection between cost and markup.

[*]College Fellow, Economics Department, Harvard University, Cambridge, MA, 02138 (email: tcunning@fas.harvard.edu). I am especially grateful to Francesco Caselli, Erik Eyster, David Laibson and Andrei Shleifer for their support and advice. I have also benefited greatly from comments by Ruchir Agarwal, Inés Moreno de Barreda, Gharad Bryan, Alex Peysakhovich, Matthew Rabin, Matan Tsur, Tomasz Strzalecki, Balasz Szentes, among many others, besides seminar participants at Harvard and LSE.

1 Introduction

The experimental literature on decision-making has found number of patterns in which decisions are influenced either by apparently irrelevant details of the choice set, or by the other choice sets presented in the experiment. One interpretation has been that these effects are due to subjects making inferences from the composition of the choice set (Wernerfelt (1995), Kamenica (2008)) but recent studies which control for these effects by randomising the composition of choice sets have confirmed many of the previously observed patterns (Ariely et al. (2003), Jahedi (2008), Mazar et al. (2009)).

The first contribution of this paper is to state a novel framework in which these effects can be described. I allow preferences to depend on the history of choice sets encountered (including concurrent choice sets), illustrated in Figure 1. In this framework choices which are made simultaneously ("joint choices") share a common history, and therefore will share common preferences. The framework thus implies that any anomalies observed in between-subjects experiments should disappear when subjects are asked to make the two decisions simultaneously. This prediction seems to be correct for a large class of anomalies.

This framework also allows for some natural distinctions between different types of history-dependence in choice. First, whether or not subjects eventually converge to consistent choice. Second, whether the convergence reflects "preference discovery" (Plott (1996)) or "preference construction" (Slovic (1995); Ariely et al. (2003)).

The second contribution of this paper is to propose a simple model of framing through what I call a *comparison effect*: I propose that being exposed to a higher value along some dimension (whether price, quality, quantity, etc.) causes subjects to put less weight on that dimension when making decisions. Put another way, when the difference between two values comes to seem *relatively* small (i.e., small relative to the other values that you are exposed to), then people will treat it as *absolutely* small, i.e. it will become less influential in decision-making. This effect can be expressed as a uniform change in the marginal rate of substitution, which has a simple formulation in terms of observable behaviour.

To apply the model of comparisons it is necessary to observe choices in which there is variation in the set of background values which serve as comparisons. An ideal domain is data collected from laboratory studies of decision-making: in a typical experiment the subjects are presented with a series of choice sets, where each choice set is composed of alternatives defined by values along well-defined dimensions. If we treat each alternative in every choice set as providing comparisons along each dimension, thus affecting preferences, the model makes a rich set of predictions. I will argue that the predictions can account for many phenomena, which cut across the traditional classifications of anomalies in choice: first, it will predict that previously encountered choice sets will affect later decisions, normally called a framing effect (Tversky and Kahneman (1981)). Second, irrelevant options in a choice set will change decisions, a phenomenon usually called menu-dependent preferences (Tversky and Simonson (1993)). Third, it predicts certain decisions normally thought to reflect the shape of the underlying utility function. In particular, the model predicts that experimental subjects will exhibit what appears to be a generalised diminishing sensitivity in their choices. Many of the predictions generated correspond to phenomena already well known under various names. The principal predictions can be summarised as follows:

- Judgments of willingness to pay (WTP) for some quantity will be affected by the quantity that was judged in previous questions. As the previous quantity increases, sensitivity to quantity falls, thus willingness to pay will fall. This association has been commonly observed in sequential judgments of willingness to pay, it is sometimes called the *contrast effect* (Day and Pinto Prades (2010)).

- Judgments of willingness to pay will also be affected by the price encountered in a previous question. As the previous price increases, sensitivity to money decreases, thus willingness to pay will rise. This association has been commonly observed in sequential choice, called the *anchoring effect* (Ariely et al. (2003)).

- Judgments of willingness to pay, chosen from a list of prices, will be affected in a similar way by the composition of the list. As the prices offered on the list get larger, judgments of willingness to pay will increase. A number of recent studies have noticed this effect, a *response scale effect* (Beauchamp et al. (2011)).

- The decision of whether to purchase a particular product will be affected by the position of unchosen "decoy" products. In one example, as the price of a decoy increases, sensitivity to money falls, so people are more likely to buy the original product. This effect is documented in the marketing literature, sometimes called a *reference price* effect (Jahedi (2008)).

- When judging willingness to pay, subjects who are asked about larger quantities are predicted to be less sensitive to incremental quantities. Thus the overall willingness to pay function will appear to have diminishing sensitivity, even if the underlying preferences (as revealed with joint choice) have no diminishing sensitivity. This phenomenon has been called *scope neglect* and *embedding* in the experimental literature (Kahneman and Frederick (2005)).

- Similarly, when two willingness to pay judgments are made simultaneously, then sensitivity to quantity will be intermediate between high and low, therefore the difference in judged value between two quantities will be strictly larger in joint evaluation than in separate evaluation. Related effects are documented in a literature on *evaluability* (Hsee and Zhang (2010)).

- When choosing between two alternatives, shifting them both out along one dimension by a fixed amount will cause decision-makers to become less sensitive to that dimension, thus causing preferences to reverse. Similar effects have been called *common difference* or *relative thinking* effects (Tversky and Kahneman (1981)).

- Finally I predict that when making demand judgments (i.e., choosing a quantity to buy at a given price), an increase in the price causes a decrease in the sensitivity to money, causing demand to be more elastic in joint evaluation than in separate evaluation.

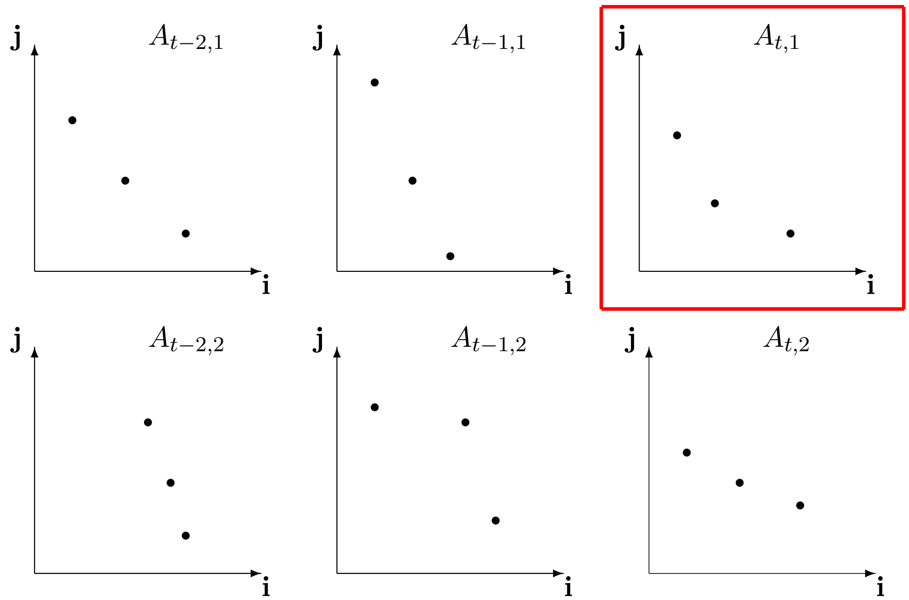


Figure 1: A history of choice sets, as is commonly offered to subjects in an experiment. Each choice set offers a set of alternatives defined on dimension i and j. The three columns represent successive phases of an experiment $(t-2, t-1, t)$, and two choice sets within each column are presented simultaneously within those phases. Many anomalies of choice can be described as the choice from set $A_{t,1}$ being systematically affected by the position of the alternatives in the whole history of choice sets.

The paper is closely related to recent work in which preferences are dependent on the choice set (Tversky and Simonson (1993); Kőszegi and Szeidl (2011); Bordalo, Gennaioli, and Shleifer (2011)). The model presented here differs in a number of respects. First, in having a different fundamental mechanism (though my comparison effect is related to the "diminishing sensitivity" in Bordalo et al. (2011)). Second, I state comparative statics with respect to the marginal rate of substitution, which means the model has very weak restrictions on the functional form of the underlying utility function. Third, most of my predictions are not contingent on observing the entire choice set, a requirement for most other models. And finally, my assumption that utility depends on the entire choice history, not just the current choice set, generates a wider range of testable predictions, and so is a candidate to explain a much wider range of phenomena.

After describing the model and the evidence I discuss a possible inference interpretation of the observed behaviour. Suppose people are uncertain about the value of a good, and when they observe larger values of that good they infer that it has a lesser marginal value.[1] This will generate a comparison effect in behaviour. There is a close analogy with recent work in cognitive science, where contrast effects (analagous to the contrast effects described above) have a natural interpretation as arising from Bayesian inference in the face of an unobserved common factor. However, both in perception and in choice, judgments remain influenced by comparisons even when the comparisons are known to be uninformative: in choice, the effect occurs when comparisons and randomised, in

[1] This is close to the arguments used in Wernerfelt (1995) and Kamenica (2008).

perception the effect occurs in optical illusions. The effect of comparisons on choice could therefore be thought of as a heuristic in the sense of Tversky and Kahneman (1974), i.e. a rule which is optimal in most circumstances, but fails to condition on all available information, and thus gives rise to systematic biases.

Moving the comparison model to real-world choices, I model how comparison effects will affect pricing in market equilibrium. If the choice set consists of the same product available at different stores then the demand system, which would ordinarily imply a constant absolute markup, now implies that higher cost will be associated with a higher markup, higher price dispersion, and more entry. I discuss existing IO evidence that is consistent with this data.

Finally I introduce a new dataset of 3,500 costs and prices from a drugstore, collected by hand. The data is unusual, because retailers are generally reluctant to release data on cost. The data reveal a number of interesting patterns in markup, but in particular they show a very tight relationship between cost and price, as predicted by the model, and which is puzzling without this model.

2 Model

In this section I begin by defining a choice history, H, and make some assumptions about how choice depends on H. I show then that a monotonic effect on the marginal rate of substitution (MRS) is equivalent to a monotonic effect on willingness to pay (WTP), and that both imply a monotonic effect on demand. When WTP is measured with a discrete choice set, MRS likewise has a monotonic effect.

I then move to comparisons between joint and separate evaluation. Because the choice history is qualitatively different between joint and separate evaluation, I require some regularity conditions on the utility functions. I am then able to derive testable predictions about the relationship between joint and separate judgment in choice, WTP, and demand.

Let the space of all possible alternatives be $\mathcal{X} \subseteq \mathbb{R}^n$, where n is the number of attributes that alternatives differ on. I will write x_i to refer to the value of $x \in \mathcal{X}$ along dimension i, and sometimes use x_{-i} to refer to a vector excluding dimension i.

The set of possible choice sets is simply $\mathcal{A} = 2^{\mathcal{X}} \backslash \{\varnothing\}$.

A choice history is a vector, with dimension $T \in \mathbb{N}$, of sets of choice sets, thus the set of possible histories is $\mathcal{H} = \left(2^{\mathcal{A}} \backslash \{\varnothing\}\right)^T$. I will refer to the elements of a choice history as $H = (H_1, .., H_T)$. The principal interpretation of a choice history is the setup of a typical choice experiment, in which a subject is exposed to T sequential stages, and thus H_t is the set of choice sets faced at time t. Typically in an experiment only one choice, chosen randomly, is implemented, so that each decision can be treated as having independent payoffs.

Observable behaviour will be a choice function $c : \mathcal{A} \times \mathcal{H} \to \mathcal{A}$, such that $c(A, H) \subseteq A$ with $A \in H_T$. Note that for simplicity the choice function is only defined for the final stage in choice history, i.e. H_T, and note that any two choice sets in H_T will share identical histories.

Assumption 1 (**Representation**). *c has a utility representation, denote it by $U(x, H)$, where*

$U(\cdot, H)$ is strictly quasiconcave, twice continuously differentiable, strictly increasing, and unbounded in every argument.

The important implication of this assumption is that choice from simultaneous choice sets will be made using the same utility function, and therefore must obey the usual restrictions on revealed preference.

I introduce two further regularity conditions, which are helpful for being able to compare results in joint and separate contexts. As the model is currently set up the utility function can change arbitrarily depending on the structure of the history, thus separate and joint choices could use entirely different utility functions. I therefore assume that the effect of history can be summarised with a sufficient statistic of that history, and that the range of that statistic is unbounded for any attribute in the history.

Assumption 2 (**Sufficient Statistic**). *There exists a function $f : \mathcal{H} \to \mathbb{R}$, monotonic, continuous and unbounded in every argument, and a choice function $d : \mathcal{A} \times \mathbb{R} \to \mathcal{A}$ with $d(A, r) \subseteq A$, such that $d(A, f(H)) = c(A, H)$ for all $A \subseteq H_T$.*

A tighter assumption is that the function depends only on the average values observed in the entire history. This assumption is needed for just one of the propositions that follow.

Assumption 3 (**Average.**). *There exists a function $g : \mathcal{H} \to \mathbb{R}$ which can be written as $g(\bar{a}(H))$, where $\bar{a}_i = \frac{1}{|C|} \sum_{c \in C} c_i$, and $C = \cup_{t=1}^{T} \cup_{A \in H_t} A$, and for all $H \in \mathcal{H}$, $g(H) = f(H)$.*

I next make the principal assumption of the model, that if any element in the history increases along dimension i, then the marginal rate of substitution between i and every other dimension will decrease.

Definition 1 (**Translation**). *For any two histories $H^A, H^B \in \mathcal{H}$, we say H^B is a translation of H^A along dimension i if there exist two alternatives $a, b \in \mathcal{X}$, with $|b_i| > |a_i| > 0$ and $b_j = a_j$ for $j \neq i$, and there exist two choice sets $A, B \in \mathcal{A}$, and a time $t \in \{1, .., T\}$, such that $H_s^A = H_s^B$ for $s \neq t$, $H_t^B = H_t^A \backslash \{A\} \cup \{B\}$, and $A = B \backslash \{a\} \cup \{b\}$.*

Assumption 4 (**Comparison Effect**). *If H^B is a translation of H^A along dimension i then for all $x \in \mathcal{X}$, $j \neq i$,*

$$MRS_{i,j}(x, H^B) \leq MRS_{i,j}(x, H^A)$$

Where we define the marginal rate of substitution as

$$MRS_{i,j}(x, H) \equiv \frac{\partial U(x, H)/\partial x_i}{\partial U(x, H)/\partial x_j}$$

Before continuing it is worth noting a limitation of this analysis: when treating WTP and Demand decisions, I will sometimes treat these as taken from a discrete choice set, and sometimes from a continuous choice set. Having a discrete choice set is an accurate representation of most experiments, and is a neat fit for my assumptions on how comparisons work. However in some of the propositions below I make predictions for how H affects a continuous choice variable. Although the *direction* of effects is generally the same between discrete and continuous choice, when I make predictions about a quantitative measure, such as a difference or an elasticity, it is very awkward to express these results while taking into account the discreteness of a subject's choices. In this paper the best I can do is conjecture that the continuous results will hold as limits of discrete choice, when the density of alternatives goes to infinity.

I now define a compensation function, i.e. a quantity of good j which would make a decision-maker indifferent between alternatives a and b. Although the definition is general, for clarity I will refer to the compensation function as a Willingness to Pay funciton.

Definition 2 (**Willingness to Pay**). *For any $a, b \in \mathcal{X}$, $j \in \{1, .., n\}$, $H \in \mathcal{H}$, define $WTP_j(b, a, H)$ as*[2]

$$U(a, H) \quad = \quad U((b_j - WTP_j(b, a, H), b_{-j}), H)$$

Because $U(\cdot, H)$ is continuous and increasing in every argument, $WTP_j(b, a, H)$ will be smoothly increasing in every b_i and smoothly decreasing in a_i for every $i \in \{1, .., n\}$. This allows us to make our first prediction.

Proposition 1 (**WTP and Comparison**). *For any two histories, $H^A, H^B \in \mathcal{H}$, and dimension $j \in \{1, .., n\}$ then*

$$MRS_{j,i}(x, H^A) > MRS_{j,i}(x, H^B), \; \forall i \neq j$$

if and only if, for all $x, y \in \mathcal{X}$, $y \geq x$,

$$WTP_j(y, x, H^A) < WTP_j(y, x, H^B)$$

Definition 3 (**Demand**). *For any $x \in \mathcal{X}$, $p \in \mathbb{R}^+$, define $D_{j,i}(x, p, H)$ as*[3]

$$D_{j,i}(x, p, H) \quad = \quad argmax_q \left\{ U((x_i + q, x_j - p \cdot q, x_{-i,j}), H) \right\}$$

Because we have assumed that the utility function is quasiconcave and continuous in all its arguments, demand will be a unique function (Mas-Colell et al. (1995), p51).

Proposition 2 (**Demand and Comparison.**). *For any $j \in \{1, .., n\}$, and any two histories, $H^A, H^B \in \mathcal{H}$,*

$$MRS_{j,i}(x, H^A) > MRS_{j,i}(x, H^B), \; \forall i \neq j$$

[2] *Define $x_{-i,j}$ as the vector x minus the entries for i and j. Note the notation is implicitly redefining the order of the elements in the utility function.*

[3] *Define $x_{-i,j}$ as the vector x minus the entries for i and j. Note the notation is implicitly redefining the order of the elements in the utility function.*

implies that for all $x \in \mathcal{X}$

$$D_{j,i}(x, p, H^A) < D_{j,i}(x, p, H^B)$$

It will be useful now to define some particular kinds of choice sets. In experiments willingness to pay is usually elicited with a set of binary decisions, i.e. a joint set of choices (e.g., as in Becker et al. (1964), and in the Multiple Price List method used in Beauchamp et al. (2011)). The subject is given a list of decisions, each between some fixed prospect $x = \{x_i, x_j, x_{-i,j}\}$ and an alternative $\{x_i + \delta_i, x_j + \delta_{j,k}, x_{-i,j}\}$, where list members are indexed by k and $\delta_{j,k}$ varies across members of the list. For clarity I will suppress the arguments $x_{-i,j}$ in what follows, use dimension labels $q = i$ and $m = j$, and use $q = \delta_q$, $p_k = \delta_{j,k}$, though the argument remains general.

Definition 4 (**Multiple Price List**). *For some $x \in \mathcal{X}$, $q \in \mathbb{R}$, $m \in \mathbb{N}$, $p \in \mathbb{R}^m$, then let $MPL(x, q, p)$ be a set of choice sets $\{(x_q, x_m), (x_q + q, x_m + p_k)\}$, for $k = 1, .., m$.*

Note that the MPL can be used to measure both willingness to pay (when $q > 0$ and p is uniformly negative), or willingness to accept (where $q < 0$ and p is uniformly positive). Define $DWTP(x, q, p, H)$ as the highest value in p for which $(x_q + q, x_m + p_k)$ is chosen over (x_q, x_m), given history H.[4]

The following corollary shows that increasing a quantity comparison will lower WTP, and increasing a price comparison will raise WTP.

Corollary 1 (**Effect on WTP**). *For any $x \in \mathcal{X}$, $q \in \mathbb{R}$, $m \in \mathbb{N}$, $p \in \mathbb{R}^m$, $H^A, H^B \in \mathcal{H}$, with $H^B_T, H^A_T \supseteq MPL(x, q, p)$, if H^B is a translation of H^A along the price dimension then $DWTP(x, q, p, H^B) \geq DWTP(x, q, p, H^A)$ and if H^B is a translation of H^A along the quantity dimension then $DWTP(x, q, p, H^B) \leq DWTP(x, q, p, H^A)$.*

A related effect occurs when the range of the MPL is manipulated. If the distribution of prices shifts out (in the sense of first-order stochastic dominance) then judged WTP will increase. The effect has to be phrased somewhat more carefully, because the WTP judgments are now chosen from different sets.

Corollary 2 (**Response Range Effect**). *For any $x \in \mathcal{X}$, $q \in \mathbb{R}^+$, and $p, p' \in \mathbb{R}^m$, if p' first-order stochastically dominates p, then for any $p_k \in p$, if $p_k \in p'$ and*

$$U((x_q, x_m), H) \leq U((x_q + q, x_m + p_k), H) \quad \Rightarrow \quad U((x_q, x_m), H') \leq U((x_q + q, x_m + p_k), H')$$

This shows that willingness to pay will be at least as high under p' than under p.

Next I define a budget set, i.e. a linear schedule of prices and quantities for some good. To represent this as a finite choice set I assume that it consists of a list of alternatives: buying zero units, buying one unit for p, buying two units for $2p$, etc., up to some maximum limit m. The limit m may seem arbitrary, but the results do not depend on the choice of m. Again I simplify to consideration of just two dimensions.

[4]Named DWTP to represent *discrete* willingness to pay, as opposed to continuous WTP, defined earlier.

Definition 5 (**Budget Set**). *For some $p \in \mathbb{R}^+$, $m \in \mathbb{N}^+$, let $B(p,m)$ be a choice set with m elements, each element $\{(0,0),(i,p\cdot i)\}$, for $i=1,..,m$.*

Corollary 3. *For any two histories, $H^A, H^B \in \mathcal{H}$, $t \in \{1,..,T\}$, if $H_s^A = H_s^B$ for $s \neq t$, and $\exists p,p' \in \mathbb{R}^+$, $p' > p$, $m \in \mathbb{N}^+$, with $B(p,m) \in H_t^A$ and $H_t^B = H_t^A \backslash \{B(p,m)\} \cup \{B(p',m)\}$, then*

$$MRS_{m,q}(x,H^A) > MRS_{m,q}(x,H^B)$$

This simply says that offering a subject a budget set with a higher price will tend to lower the subject's sensitivity to money. It follows very simply, because $B(p',m)$ can be derived, by a set of translations, from $B(p,m)$.

3 Predictions and Evidence

The predictions can be divided into two parts: First, evidence for the general framework I have been using (in which the choice function has a utility representation, conditional on the choice history H). Second, for the predictions of the assumption about comparative judgment.

3.1 Evidence for the Framework

Assumption 1 states that choice, conditional on the history H, has a utility representation. This implies that choices with the same history (i.e. choices taken jointly) will satisfy the strong axiom of revealed preference (Richter (1966)), i.e. there will be no cycle of alternatives $x^1, x^2, ..x^n$ such that each element is revealed to be preferred to the next, and x^n is revealed preferred to x^1.[5] I will refer to violations of SARP as "intransitive" choices.

What evidence is there that subjects are consistent when they make choices jointly? Unfortunately there are very few studies which directly address this question. Tversky and Kahneman (1986) say that "the major finding of the present article is that the axioms of rational choice are generally satisfied in transparent situations and often violated in non-transparent ones," though they report almost exclusively experiments in which the axioms are violated, taken from between-subject studies.

The simplest test would take evidence of intransitive choices from a between-subjects study, e.g. where subjects choose x from $\{x,y,z\}$ and y from $\{x,y\}$, and present subjects with the two choice sets jointly, to see if the inconsistency survives. Closest to this is Mazar et al. (2009), who find that judgments of WTP differ depending on the set of options given for their WTP judgment. They then show that the discrepancy almost disappears when the two judgments are made side by side.

Similarly Ordóñez et al. (1995) find that classic preference reversals (in which the WTA for a high-prize lottery is higher than for a high-probability lottery, though the latter would be chosen

[5] Note that the weak axiom of revealed preference (WARP) is not strong enough: a cycle of intransitive binary choices, $x \succ y \succ z \succ x$, satisfies WARP, but does not have a utility representation.

over the former, Lichtenstein and Slovic (1971)) decreased significantly when presented jointly, and were almost eliminated when the game was played for real stakes.

Some evidence contrary to this prediction comes from experiments in which subjects choose strictly dominated options from a single choice set, as in Birnbaum et al. (1999), or a strictly dominated pair of options from a pair of choice sets, both of which are implemented (Tversky and Kahneman (1986) and Rabin and Weizsacker (2009)). However in these cases the alternatives are quite complicated to understand, Tversky and Kahneman report that "people who are given these problems are very surprised to learn that the combination of two preferences that they considered quite reasonable led them to select a dominated option".

Of related interest are studies in which two framed choices are presented side by side, to discover whether the framing effect survives. Hsee (1998) gives examples in which irrelevant attributes affect the WTP for certain objects (e.g. the WTP for a given amount of ice cream changes depending on the size of the cup) yet when judging payments side by side, willingness to pay is independent of cup size.

It is worth mentioning that, if subjects infer payoff-relevant information from the choice set (as in Wernerfelt (1995); Kamenica (2008)), then we would expect the same relationship between history and transitivity: because joint choices share the same history, they would be made with the same information, thus choices should obey transitivity.

The following subsections discuss evidence for the predictions of the comparison effect.

3.2 Influence Between Choice Sets

Proposition 1 predicts that WTP will decrease if the choice history is translated upwards along the quantity dimension. This is a regular finding in the contingent valuation literature, Day and Pinto Prades (2010) say:

> The evidence that exists, therefore, conforms with a commodity-sequence ordering anomaly in which a good is regarded more favourably when preceded by a question offering a relatively smaller level of provision (an improving commodity sequence) while being regarded less favourably if preceded by a relatively larger level of provision (a worsening commodity sequence).

Similar sequence effects have also been found in the field, Bhargava (2008) calls it a *contrast effect* and gives example from marking exams, judging cases, or meeting potential dates, in each case evaluation of the present case tends to be lower if the preceding stimulus was higher.

Although it cannot be represented in the framework of this paper, a similar pattern seems to also hold in framing effects, in which irrelevant information (but not part of the choice history, H), affects judgment. Bartels (2006) summarises a literature on "denominator effects", in which the willingness to pay for some good is affected by the size of the pool from which it is drawn. For example, in one case the WTP to save 10 lives from a disease was much higher when subjects are told 20 lives are at risk, than when they are told 100 lives are at risk. Similarly, Hsee (1998) reports that the WTP for a given amount of ice cream is much lower when he increased the size of the container in which it is placed.

Proposition 1 also predicts that WTP for a good will increase if the choice history is translated upwards along the price dimension. Best-known is the *anchoring effect*, where the subject is first asked whether they would pay $\$p$ for a good (i.e., given the choice set $\{(0,0),(Q,p)\}$), and subsequently asked for their willingness to pay. The literature is surveyed in Chapman and Johnson (2002). In a representative example from Fudenberg et al. (2010), incentive-compatible WTP judgments for a computer keyboard were on average \$38.85 for a group with anchors between \$0 and \$20 (here the "anchor" corresponds to $\$p$), and \$47.28 for a group with anchors between \$80 and \$100.

A similar effect is observed in contingent valuation studies (summarised in Day and Pinto Prades (2010)): in these studies subjects are asked a sequence of two questions about whether they would pay a certain amount of money for some good. A higher price in the first question is associated with an increased likelihood of accepting the offer in the second question, all else equal.[6]

In marketing there is a large literature on the "reference price" of a product, where this price is usually identified with an average of previously observed prices. A recent survey of evidence (Mazumdar et al. (2005)) concludes that a higher reference price increases both probability of purchase, and quantity purchased.

The model predicts two, perhaps surprising, variations of the anchoring effect. First, because an increase in an historical price lowers overall sensitivity to money, anchoring should work across products. That is, WTP for product A should be affected by previously encountered decisions involving money, whether the decision regards product A or product B. Secondly, the model predicts that if the subject is presented with multiple anchors, the comparative statics of each of the anchors is the same.

A recent paper confirms both of these predictions: Mochon and Frederick (2011) find, first, that WTP for a product is affected by a prior anchoring question, whether the prior question is about the same product or a different product.[7] Second, Mochon and Frederick find that when presented with multiple anchors, the magnitude of both prior anchors affects subsequent WTP judgments in the same way. Both results are difficult to reconcile with standard accounts of anchoring, and to account for these effects Frederick and Mochon (2011) propose a "scale distortion" theory of anchoring.

3.3 Influence from Within the Choice Set

The predictions for effects from within the choice set regard the effect of changes in the attributes of a "decoy" alternative, i.e. an unchosen alternative. If the price of the decoy increases, customers should substitute towards the higher price goods (or should should substitute towards purchasing).

---

[6]The protocol here is choice, rather than matching, but the same comparative statics will hold. Day and Pinto Prades also mention that they find an asymmetry: the effect is strong when the first price is below the second price, perhaps because subjects have a particular aversion to paying above the reference price.

[7]Mochon and Frederick find t-statistics of 3.9 for within-product anchors, and 2.9 for between-product anchors.

If the quantity of the decoy increases, customers are predicted to substitute towards the lower quantity good (or should substitute away from purchasing).

Corollary 4 (**Decoy Effect**). *For any two histories $H^A, H^B \in \mathcal{H}$, if there exists a dimension i and two alternatives $a, b \in \mathcal{X}$, with $|b_i| > |a_i|$ and $b_j = a_j$ for $j \neq i$, and two choice sets $A, B \in \mathcal{A}$, $H_T^B = H_T^A \backslash \{A\} \cup \{B\}$, and $A = B \backslash \{a\} \cup \{b\}$, and $H_s^A = H_s^B$ for $s \neq T$, and $a \notin c(A, H^A)$, $b \notin c(B, H^B)$, then for any $x, y \in A$, with $x_i > y_i$ and $x_{-i} \leq y_{-i}$,*

$$x \notin c(A, H^A), y \in c(A, H^A) \Rightarrow x \notin c(B, H^B)$$

Most published evidence on decoy effects concerns the effect of adding an alternative to the choice set, rather than changing the location of an alternative (e.g. the "compromise" (Simonson (1989)) and "asymmetric dominance" effects (Huber et al. (1982))). The model of comparison given here does not given an unambiguous prediction in these cases, because the number of elements in H changes, so there cannot be a translation relationship. Intuitively the reason is that, without more information, we cannot say whether the new comparator is relatively high or relatively low. For example suppose you are considering whether to buy an object for \$1, and your sensitivity to price depends inversely on the average price observed. Adding an inferior object at \$1.10 may be thought to make the object seem cheap by contrast. However, if in your prior experience the item had been priced at \$2, then the addition of a \$1.10 price could lower the average price in your set of comparisons, thus weakening the sense that \$1 is a low price.

The only paper I know of which both randomises the choice set and varies the attributes of a decoy isJahedi (2008). Jahedi offers subjects the chance to buy one of two goods (or neither), and compares choices when the price of one of the goods is explicitly drawn from a random distribution. He finds robust evidence that increasing the price of the low-q alternative increases the probability of choosing the high-q alternative, even when controlling for substitution effects. For example, increasing the price of a \$5 Starbucks voucher from \$4.00 to \$4.40 increased the probability of buying 2 Starbucks vouchers (priced at \$4.20) from 39% to 47% (and in both cases, less than 0.5% of the subjects purchased the single voucher, i.e the low-q decoy).

I do not know of a similar experiment which varies the quality or quantity of decoy alternatives.

3.4 Binary Choice

The remainder of the propositions compare choice between joint and separate evaluation, e.g. for some pair of choice sets A, B, we will be comparing choices when the history H_T is equal either to $\{A\}$, $\{B\}$, or $\{A, B\}$. Although we can derive predictions when $\{B\}$ is a translation of $\{A\}$, we cannot directly compare either with $\{A, B\}$ because they have a different cardinality. Thus for all the following propositions I will be adopting Assumption 2, that the effect of history on choice can be summarised with a sufficient statistic. This can be thought of as assuming that the utility function always has the same basic structure, and that choice history can affect only one parameter of calibration.

For binary choice the model predicts that adding a common factor to both alternatives along some dimension will lessen sensitivity to that dimension. Then, if decisions are unaffected by a common difference in joint choice, the common difference will cause preference reversals when the two decisions are made separately. The proposition simply states that reversals will only occur in one direction (they can only be in favour of the loser on dimension i, the dimension on which the difference is added.

Proposition 3 (**Common Difference**). *for some dimensions $i, j \in \{1, .., n\}$, and for all $x^A, x^B \in \mathcal{X}$ where $x_i^A < x_i^B$, $x_j^A > x_j^B$, and $x_k^A = x_k^B$ for $k \neq i, j$, and for $\delta \in \mathbb{R}$, and $sgn(\delta) = sgn(x_i^A) = sgn(x_i^B)$,*

$$\begin{aligned} (x_i^A, x_{-i}^A) \in c(A^L, H) &\iff (x_i^A + \delta, x_{-i}^A) \in c(A^H, H) \\ \textit{with } A^L, A^H &\subseteq H_T \\ A^L = \{(x_i^A, x_{-i}^A), (x_i^B, x_{-i}^B)\} \quad &\textit{and} \quad A^H = \{(x_i^A + \delta, x_{-i}^A), (x_i^B + \delta, x_{-i}^B)\} \end{aligned}$$

then for any $H^L, H^H \in \mathcal{H}$, with $A^L \in H_T^L$, $H_T^H = H_T^L \backslash \{A^L\} \cup \{A^H\}$ and $H_s^H = H_s^L$, $\forall s \neq T$,

$$(x_i^A, x_{-i}^A) \in c(A^L, H^L) \quad \Rightarrow \quad (x_i^A + \delta, x_{-i}^A) \in c(A^H, H^H)$$

Common difference effects are well known in decisions about purchasing goods. Most famous is an example from Tversky and Kahneman (1981): they found that a majority of subjects were willing to walk 20 minutes to buy a good at $10 instead of $15, but in a separate experiment a majority were *not* willing to walk 20 minutes to buy a good at $120 instead of $125. The decisions are surprising because both decisions seem to offer the same tradeoff, between 20 minutes of time and five dollars. This example fits the pattern just described, in which adding a common difference reverses preferences.

To illustrate the magnitude of these responses, Azar (2011) reports that subjects were asked how large a saving would justify a 20 minute trip to another store. When contemplating buying a $10 pen, the median response was a $4 saving. When contemplating a $1000 computer, the median response was a $50 saving. When contemplating a $10,000 car, the median response was a $300 saving.[8] A limitation of this literature has been that all published experiments have asked subjects only hypothetical questions, excepting one field experiment with bagels and cream cheese, documented in Azar (2009), which failed to reproduce the effect.

Tversky and Kahneman (1981) explain their result as due to diminishing sensitivity in the valuation of changes in wealth, evaluated using a concave function $v(\cdot)$, "[B]y the curvature of v, a discount of $5 has a greater impact when the price of the calculator is low than when it is high." It is worth noting that this explanation implicitly sets a reference point: the reference point is at paying nothing, i.e. at not purchasing the product. This is in contrast to most applications of prospect

[8] This also predicts an effect of price on add-on purchases, as Savage (1954) puts it: "a man buying a car for $2,134.56 is tempted to order it with a radio installed, which will bring the total price to $2,228.41, feeling that the difference is trifling. But, when he reflects that, if he already had the car, he certainly would not spend $93.85 for a radio for it, he realizes that he made an error."

theory to purchase behaviour, where the reference point is set at the purchase of the good, and changes are evaluated relative to that point (e.g. in Kőszegi and Rabin (2006) the reference point is the rationally expected outcome; in this situation it would seem to be purchase at one store or other). If the reference point is purchase, then the comparison of gain-loss utility is a comparison between $v(0)$ and $v(5)$[9], which does not change in the two scenarios, and thus would not predict the effect.

The model has a clear prediction: that the diminishing sensitivity will disappear in joint choice. Normatively, this has a strong appeal: because in both situations you are making the same tradeoff between time and money. Unfortunately I do not know of any published experiments in which both problems are faced jointly.

Common difference effects also occur in other domains, usually interpreted as evidence for diminishing sensitivity, and comparison effects may also be involved in these other cases. In intertemporal choice, if two delayed rewards are offered (one smaller and sooner, one larger and later), then preferences are often reversed in favour of the larger-later reward when both rewards are pushed back in time by a common difference (Green et al. (1994), and see Frederick et al. (2002) for a survey). In risky choice the common difference effect is sometimes identified with the Allais paradox (Allais (1953)): increasing the probability of a prize common to all gambles by a fixed amount should not change preferences between those gambles, according to expected utility, but it does, in the same direction as the other examples. Intuitively the effect is also seen in Zeckhauser's paradox: if playing Russian roulette, it seems to many people more valuable to remove the last and only bullet from the gun, than to reduce the number of bullets from 4 to 3, although both actions decrease your chance of death by 1/6. It would be interesting to see these experiments run as joint choices.

3.5 Willingness to Pay

The model predicts that if a WTP function has a fixed elasticity in joint evaluation, then it will have a lower elasticity in separate evaluation. Intuitively as the quantity increases, WTP increases, but the comparison effect has a dampening effect on the increase, because the higher quantity lowers sensitivity. Thus the overall WTP function will show steeply diminishing sensitivity (i.e., it will be inelastic). However under joint evaluation the comparison effect will disappear.

In the following I use a simplified willingness to pay function, $WTP(q, H)$, short for $WTP_j((x_i + q, x_{-i}), x, H)$.

Proposition 4 (**Scope Neglect**). *If there exists an* $\varepsilon \in \mathbb{R}$ *such that for any* $x \in \mathcal{X}$, $q_L, q_H \in \mathbb{R}_+$, $m \in \mathbb{N}$, $p \in \mathbb{R}^m$,

$$\frac{WTP(q_H, H) - WTP(q_L, H)}{WTP(q_L, H)} \frac{q_L}{q_H - q_L} = \varepsilon$$

where $MPL(x, q_L, p), MPL(x, q_H, p) \in H_T$, *then*

$$\frac{WTP(q_H, H^H) - WTP(q_L, H^L)}{WTP(q_L, H^L)} \frac{q_L}{q_H - q_L} < \varepsilon$$

[9] Or $v(-5)$ and $v(0)$, if purchase from the other store is the reference point.

where $MPL(x, q_H, p) \subseteq H_T^H$, $H_T^L = H_T^H \backslash \{MPL(x, q_H, p)\} \cup \{MPL(x, q_L, p)\}$, *and* $H_s^L = H_s^H$ *for* $s \neq T$.[10]

There is a great deal of evidence that WTP functions are often surprisingly inelastic in separate evaluation. Kahneman and Frederick (2002) cite a range of evidence for what they describe as "scope neglect" in judgments of willingness to pay, where subjects report similar judgments of WTP for very different magnitudes of some good.[11] In the best-known example (from Desvousges et al. (1993)) subjects were asked their willingness to pay to save birds from death in oil ponds. The mean WTPs for saving 2,000, 20,000, or 200,000 birds were \$80, \$78, and \$88, respectively.[12] [13]

Before going on it is important to note that there is an error in the usual description of these effects. They are often called anomalies of "insensitivity" or "neglect", and explained as anomalous because WTP judgments are either entirely insensitive to x, or there is "nearly complete" neglect of scope, i.e. the slope of the line is very small.[14] However a better description would be that the evidence shows *low elasticity* in willingness to pay. It is not anomalous for Desvousges' subjects to value an incremental 198,000 birds at \$8 (i.e., a low slope of the WTP function). It is anomalous for a 100-times increase in the number of birds to only evoke a 10% increase in willingness to pay (i.e., a very low elasticity).

Finally, the theory here predicts that in joint choice, the WTP function should have a much higher elasticity. This is what is found by Hsee and Zhang (2010), in comparing WTP for chocolates, judged jointly and separately. Although using sequential choice, not joint, Kahneman and Frederick (2002) also report that subjects show much greater elasticity with respect to the x variable in within-subjects experiments than in between-subjects experiments.

3.6 Demand

The argument about WTP transposes very naturally to demand: when the price of a good increases, demand will fall, but the fall will be dampened through the comparison effect lowering sensitivity to money. Thus if demand has a fixed elasticity in joint evaluation, then it should be less elastic in separate evaluation.

Proposition 5 (**Demand Inelasticity**). *If there exists an* $\varepsilon \in \mathbb{R}$ *such that for any* $x \in \mathcal{X}$,

[10] Suppressing the arguments x and p, common to all the $WTP(\cdot)$ functions.

[11] See also discussion in Frederick and Fischhoff (1998)

[12] Carson (1997) argues against the existence of scope neglect, but only going so far as saying that there is *some* response to scope. For example, he cites a study which finds an average WTP of \$3.78 to prevent 0.04 deaths per 100,000 per year, and \$15.23 to prevent 243 deaths per 100,000 per year. This shows that there is not complete neglect of scope, yet the function is very far from linear (the implied value of a statistical life is \$9.5 million in the former case, and \$6,300 in the latter).

[13] Scope neglect is also sometimes known as the "embedding effect", Diamond and Hausman (1994).

[14] Kahneman et al. (1999)define insensitivity to scope as "the quantitative attribute has little weight in the valuation. "

$p_L, p_H \in \mathbb{R}^+$, $m \in \mathbb{N}$, $q \in \mathbb{R}^m$

$$\left| \frac{D_{j,i}(p_H, H) - D_{j,i}(p_L, H)}{D_{j,i}(p_L, H)} \frac{p_L}{p_H - p_L} \right| = \varepsilon$$

where $B(p_L, q), B(p_H, q) \subseteq H_T$, *then*

$$\left| \frac{D_{j,i}(p_H, H^H) - D_{j,i}(p_L, H^L)}{D_{j,i}(p_L, H^L)} \frac{p_L}{p_H - p_L} \right| < \varepsilon$$

where $H_T^L = H_T^H \backslash \{B(p_L, q)\} \cup \{B(p_H, q)\}$, *and* $H_s^L = H_s^H$ *for* $s \neq T$.[15]

An ideal experiment would be to elicit conditional demand functions, e.g. by giving subjects a list of prices, so they could choose quantities they would buy at a given price level. Elasticities elicited in this way are predicted to be lower than elasticities elicited through asking for demand levels separately. Unfortunately I am not aware of any studies which use this protocol.

There is an interesting ancillary prediction which this makes. Suppose that the recent price of a product serves as a salient comparator, then elasticity should be larger in the short-run than in the long-run (the opposite to the Le Chatelier principle). When the price goes up, the short-run effect will be a large drop in demand, because the new price seems large compared to the current price. As the comparison fades, demand will bounce back. An analogous effect holds for decreases in price.

There are some other papers which make related predictions to mine. Mazar et al. (2009) observe that WTP tends to increase with some measure of the reference price. They treat this effect as a distortion of true preferences, which allows them to draw two demand curves: a "true" demand curve, and an observed demand curve. Mazar et al. show that the observed demand curve will be less elastic than the true demand curve. However that paper does not give any method for estimating the true demand curve. The model presented here does not specify a "true" demand curve, but it does give the testable prediction that demand elasticity will be greater in joint than in separate evaluation.

Hsee and Zhang (2010) have a theory of evaluability, proposing that subjects are generally more sensitive to differences in joint than in separate evaluation, and so predicts that demand elasticity will be higher in joint evaluation than in separate evaluation. Also, Ariely et al. (2003) make a similar prediction, that elasticity may be larger in the short-run than in the long-run.

3.7 Joint-Separate Reversals in WTP Ranking

If the marginal rate of substitution depends just on the average value observed along some dimension (as in Assumption 3) then it becomes simple to compare joint and separate modes of evaluation, the sensitivity in joint evaluation will always be intermediate between the sensitivities in separate

[15] Suppressing the argument m, common to all the $B(\cdot)$ and $D(\cdot)$ functions.

evaluation. The difference in WTP will thus be strictly larger in join than in separate evaluation. As above, I will use an abbreviated willingness to pay function, $WTP(q, H)$.

Proposition 6 (**WTP Reversal**). *If choice satisfies Assumption 3 then, for some* $q_H > q_L$*,*

$$WTP(q_H, H) - WTP(q_L, H) > WTP(q_H, H_H) - WTP(q_L, H_L)$$

where $MPL(x, q_L, p), MPL(x, q_H, p) \in H_T$*, and* $H_T^L = H_T^H \backslash \{MPL(x, q_H, p)\} \cup \{MPL(x, q_L, p)\}$*, and* $H_s^L = H_s^H$ *for* $s \neq T$*.*

Hsee et al. (1999) survey a literature of reversals between joint and separate modes of evaluation, often described as finding that people are more sensitive to a given difference in joint evaluation than in separate evaluation. However the evidence usually is not through comparing the size of differences in WTP (as in the proposition). Instead, most experiments use alternatives defined on two dimensions, and find reversals of *ordering* in WTP. For example, when judging WTP for a 10,000-entry dictionary with a clean cover, and a 20,000-entry dictionary with a torn cover, the former has a higher WTP when judged separately, but the latter has a higher WTP when the two objects are judged jointly.

This reversal of ranking could occur in the comparison model if one of the dimensions of choice was not subject to the comparison effect (or much less subject to it). Hsee et al. (1999) discuss the properties of dimensions which are more likely to exhibit this reversal: they pick out, in particular, dimensions on which either there is limited information about the value of different objects, or otherwise difficult to evaluate, or which appeal to visceral wants.

For the dictionary example, learning from the choice set may well be a rational justification for the reversal cited: learning that there exists a 20,000-entry dictionary may tell you that the 10,000-entry dictionary will not fulfill your needs.

4 Inference Interpretation

In this section I will first show under what conditions a relative thinking effect would arise from an optimal inference. The model has a close resemblance to recent models from cognitive science, modeling perception as Bayesian inference. I then discuss how optimal inference can be reconciled with people making systematic mistakes.

4.1 An Inference Interpretation of Relative Thinking

To introduce the inference interpretation, I briefly discuss perception. The contrast effect in choice, discussed above, corresponds to a robust finding in perceptual judgment. William James says, in his *Principles of Psychology* (1890)[16]

> *A bright object appears still brighter when its surroundings are darker than itself, and darker when they are brighter than itself.*

[16] James (1890)

In psychology this is also called a contrast effect: the judgment of objective magnitude is negatively affected by the magnitude of a comparator. James described contrast effects for judgment of brightness, of hue, and of temperature, and he claimed that the contrast effect works both when a comparator is perceived at the same time as the target object (*simultaneous contrast*), or just prior (*sequential contrast*). Subsequent literature has documented contrast effects using formal experiments, finding them across many domains of perception and judgment.[17]

For much of the 20th century psychological work on anomalies in perception was concerned with giving neurological explanations, describing mechanisms without explaining why people should be subject to biases which seem easily fixed, whereas a recent literature instead interprets illusions as by-products of optimal inference from incomplete information (Kersten et al. (2004)). The visual information arriving on a retina is necessarily too weak to license any strict deductions about what is being seen; instead, inferences must be made using prior beliefs.

A common form of optical illusion is to show two identical stimuli, each surrounded by different backgrounds (Figure 2). The backgrounds differ either in brightness (lightness contrast illusion), in hue (colour contrast), in size (Ebbinghaus illusion), or in contrast (Chubb illusion). In each case the two central stimuli, although identical, appear to be different, such that each one's relative difference against their background appears to be an absolute difference. For example, the stimulus which is against a dark background seems lighter, the stimulus which is surrounded with large shapes seems smaller, etc.

There is a very elegant Bayesian inference explanation for each of these illusions (Kersten et al. (2004)). Suppose there are two stimuli, x and y, of unknown magnitude, and the subject receives noisy signals $\hat{x}$ and $\hat{y}$, and then forms judgments $E[x|\hat{x},\hat{y}]$ and $E[y|\hat{x},\hat{y}]$. The illusions can be thought of as showing that $\frac{dE[x|\hat{x},\hat{y}]}{dy} < 0$, i.e. judgment of the magnitude of one stimulus is decreasing in the magnitude of a neighbour. This will in fact be a rational inference if, for example, the signals share a common multiplicative error,

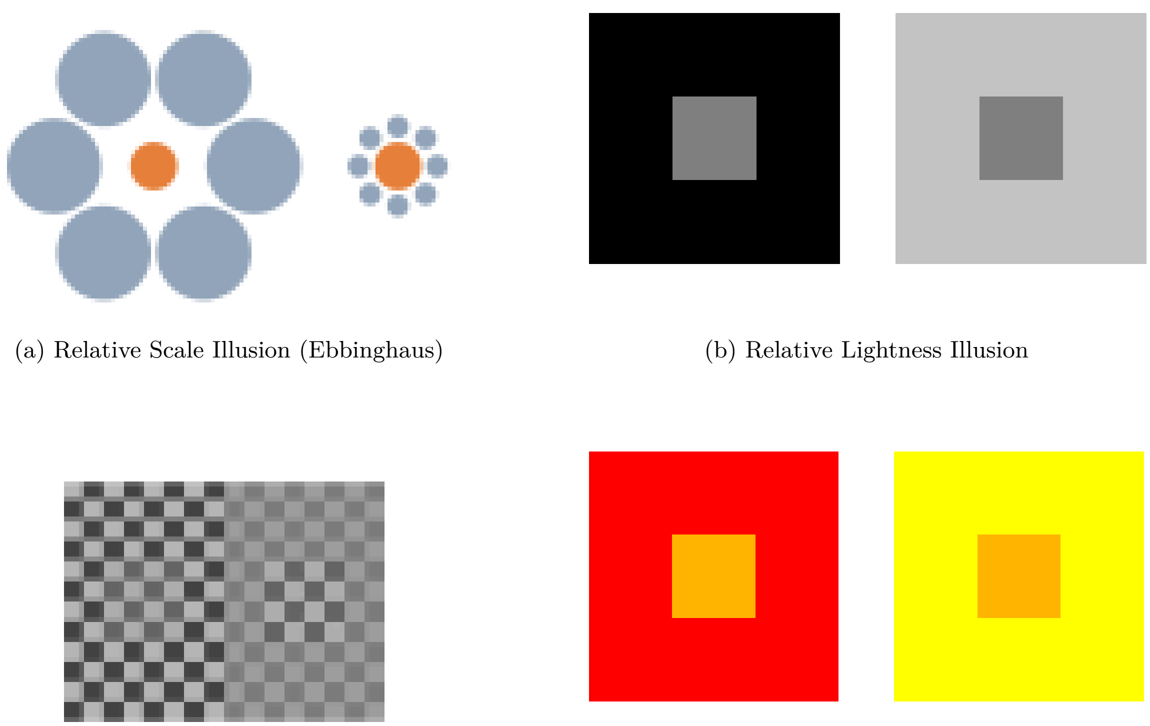
$$\begin{aligned}\hat{x} &= x/\beta \\ \hat{y} &= y/\beta\end{aligned}$$

where x, y and β are distributed such that the monotone likelihood ratio applies for the pairs $\{\hat{x}, \beta\}$ and $\{\hat{y}, \beta\}$, so that $\frac{dE[\beta|\hat{y}]}{d\hat{y}} < 0$. Thus the expectation will become

$$\frac{dE[x|\hat{x},\hat{y}]}{d\hat{y}} = \hat{x}\frac{dE[\beta|\hat{x},\hat{y}]}{d\hat{y}} < 0$$

Thus the illusions show in Figure 2 can each be explained if perception of the attribute (lightness, hue, size, contrast) is influenced by some unobserved noise. Consider subfigure (b), an illusion

[17] See Parducci (1995) for a review of literature on simultaneous contrast. Regarding sequential contrast Stewart et al. (2006) say, surveying results from across many domains of perception, "the details of the results depend somewhat on the experimental task but the general pattern is robust: The response on the current trial is biased [...] away from the stimulus presented on [the previous] trial." Note also that the effect of interest is *objective* judgment, e.g. which of two cards is lighter, or how long is a line in centimetres.



(a) Relative Scale Illusion (Ebbinghaus)

(b) Relative Lightness Illusion

(c) Relative Contrast Illusion (Chubb)

(d) Color Contrast Illusion

Figure 2: In each case the left-hand and right-hand central patches are identical. Most observers find that they appear dissimilar, in the opposite way that the backgrounds are dissimilar.

of differing lightness. An unobserved common factor in perception of lightness of an object (or its reflectivity) is the brightness of light incident upon the scene. As you receive more light from an object you revise upwards both your estimate of the reflectivity of the object, and your estimate of the incident light. If the adjacent objects are known to receive the same incident light, then increasing your estimate of the incident light will cause you to decrease your estimate of the reflectivity of the original object. In this way a light background can cause observers to judge an object to be darker, even perfectly rational observers.

In the other optical illusions there are similar explanations through unobserved variables. For an illusion of hue, it can come from unobserved hue of the incident light. For an illusion of size, it can come from an unobserved distance to the distance of objects. For contrast, it can come from unobserved haziness of the air.

To summarise, existence of a common unobserved factor can explain contrast effects, because it implies that relative observed magnitude is a good proxy for absolute magnitude. Transposing this idea from judgment to choice, there exists a simple model of inference that will give rise to relative thinking. The technical difference is just that in choice we usually treat the magnitudes (x) as known, but the importance of each attribute (β) is unknown. The following are sufficient conditions for relative thinking:

Definition 6. A decision-maker is described as having **monotonically unknown separable preferences** if
(i) they have a utility function $U(x_1, ..., x_n) = E\left[\sum_{i=1}^{n} \frac{f_i(x_i)}{h_i(\beta_i)}\right]$ with f_i and h_i both smooth and strictly increasing functions,
(ii) the coefficients $\beta_1, .., \beta_n$ are unknown to the decision-maker (but the decision-maker has smoother priors over each β)
(iii) the DM believes each comparator $c_i \in C$ (which includes all the alternatives in A) to be iid draws from a distribution $G_i(c_i|\beta_i)$, which satisfies the monotonic likelihood ratio with respect to c_i and β_i[18].

Proposition 7. *If the DM can be described as having monotonically unknown separable preferences then*

$$\frac{d}{dc_i^k} \frac{\partial E[U(x)|C]/\partial x_i}{\partial E[U(x)|C]/\partial x_j} < 0$$

Note that the decision-maker must take the expectation of the utility function, because the βs are uncertain. Thus proposition 7 shows the effect of a comparator on the marginal rate of substitution, reproducing the relative thinking effect.

An agent with preferences of this kind will thus rationally exhibit all the effects I have discussed as anomalies: contrast, anchoring, scope neglect, common difference, etc.

A simple parametric example can be given, restricting comparators to just the choice set $\{x^1, .., x^n\}$. Suppose $U(x) = E \sum_i \frac{x_i}{\beta_i}$, with $ln(x_i) \sim N(\beta_i, \sigma_i^2)$ and priors over the β's such that

[18] I.e. $\frac{g_i(R',\beta')}{g_i(R,\beta')} > \frac{g_i(R',\beta)}{g_i(R,\beta)}$ for $R' > R$, $\beta' > \beta$.

$ln(\beta_i) \sim N(\mu_i, s_i^2)$. Then utility will be

$$
\begin{aligned}
E[U(x)] &= \sum_i x_i E[\beta_i^{-1}] \\
&= \sum_i x_i exp\{-E[ln(\beta_i)] - \frac{1}{2}V[ln(\beta_i)]\} \\
&= \sum_i x_i exp\left\{-\frac{s_i^{-2}\mu_i + \sigma_i^{-2}\sum_j ln(x_i^j)}{s_i^{-2} + n\sigma_x^{-2}} - \frac{1}{2}(s_i^{-2} + n\sigma_i^{-2})\right\} \\
&= \sum_i x_i \left(\prod_j x_i^j\right)^{-\gamma_i} \phi_i
\end{aligned}
$$

where

$$
\begin{aligned}
\gamma_i &= \frac{\sigma_i^{-2}}{s_i^{-2} + n\sigma_x^{-2}} \\
\phi_i &= exp\left\{-\frac{s_i^{-2}\mu_i}{s_i^{-2} + n\sigma_x^{-2}} - \frac{1}{2}(s_i^{-2} + n\sigma_i^{-2})\right\}
\end{aligned}
$$

Thus $E[\beta_i]$ is a decreasing function of the geometric mean of the values of x observed along dimension i (holding n fixed). Because the utility function satisfies the condition on the marginal rate of substitution, it will satisfy all the properties given in the previous section. I quickly demonstrate two. First, because utility is linear in money, as in all goods, using it as numeraire we can express willingness to pay for an increment τ_i as $WTP(\tau_i) = E[U(\tau_i)]$ (suppressing the other arguments in the utility function). Then, in this model, the WTP function will have elasticity of exactly 1 in joint evaluation. But in separate evaluation, the function will have a decreasing slope:

$$
\begin{aligned}
\frac{dWTP(\tau_i)}{d\tau_i} &= \frac{dE[U(\tau_i)]}{d\tau_i} \\
&= E[\beta_i^{-1}] - \frac{\sigma_i^{-2}}{s_i^{-2} + n\sigma_i^{-2}} E[\beta_i^{-1}] \\
&< E[\beta_i^{-1}]
\end{aligned}
$$

Thus the elasticity is less than one:

$$
\begin{aligned}
\frac{dWTP(\tau_i)}{d\tau_i}\frac{\tau_i}{WTP(\tau_i)} &= \frac{dE[U(\tau_i)]}{d\tau_i}\frac{\tau_i}{WTP(\tau_i)} \\
&= \frac{dE[U(\tau_i)]}{d\tau_i} E[\beta_i^{-1}] \\
&= 1 - \frac{\sigma_i^{-2}}{s_i^{-2} + n\sigma_i^{-2}} < 1
\end{aligned}
$$

Second, I find that there is a contrast effect: an increase in the magnitude of one member of the choice set (τ_i^b) will decrease the WTP for all other members of the choice set.

$$\frac{dWTP(\tau_i)}{d\tau_i^b} = -\frac{\sigma_i^{-2}}{s_i^{-2} + n\sigma_i^{-2}} E[\beta_i^{-1}] < 0$$

4.2 Is the Inference Conscious?

The explanation for relative thinking just given treats it as optimal inference: when there is an unobserved common factor, then the *relative* magnitude of an object is a good proxy for its *absolute* value. This theory also fits the fact that menu effects disappear in joint choice: because in joint choice both decisions condition on the same information set.

There is a small literature on information-based explanations of violations of revealed preference. Sen (1993) discusses examples where it is due to the "epistemic value of the menu". Wernerfelt (1995) and Kamenica (2008) both give equilibrium rationalisations of a decoy effect in the purchase of goods; where, for example, you may rationally choose a large hat from the choice set $\{medium, large\}$, but a medium hat from $\{small, medium, large\}$, because the composition of the choice sets conveys payoff-relevant information that the firm knows themselves. Prelec et al. (1997) show in experiments that inference can explain a large part of certain previously observed decoy effects.

However, although inference from the choice set is surely important, relative thinking appears to remain even when information effects are controlled for. First, choice set effects occur even in choice among monetary gambles, where it would be a stretch of the usual interpretation of choice to say that consumers infer something about the value of money.[19] Second, frames and irrelevant alternatives continue to affect choice even when explicitly randomised. For example Ariely et al. (2003) shows the anchoring effect works with digits from subjects' social security numbers, and Jahedi (2008) finds decoy effects with randomly generated choice sets, and contrast effects occur much more strongly between consecutive elements in sequential choice, than between distant elements (Bhargava (2008)), whereas if the elements are generated by an i.i.d. process the recentness of a comparison should be irrelevant.

Again there is an analogy with perception: optical illusions occur when people apply a rational heuristic, but fail to integrate extra knowledge. In a contrast illusion, contrast with the background is usually a good cue in judging the shade of an object. However in a typical optical illusion the illumination is clearly uniform over the whole area, so local contrast is not informative (because illumination does not vary at that level). Similarly, if the Ebbinghaus illusion is based on a rule which uses relative size as a proxy for distance, this rule should not be used in this case, because it is clear that both inner circles are the same distance away, thus their size relative to surrounding circles is irrelevant. Despite this, the illusion persists.

Similarly, people may use a relative position heuristic in choice, even when they are aware that an object's relative position is uninformative. Overall this seems to closely fit the program of heuristics and biases outlined in Tversky and Kahneman (1974) and elaborated in Kahneman and

[19] See Birnbaum (1992), Herne (1999), and Stewart et al. (2003)

Frederick (2005), in which simple rules are used which perform well in most typical contexts, and biases occur when those rules are applied outside that set of contexts.

5 Discussion

5.1 Structure on Choice History

There are some other remarks worth making with respect to the framework, i.e. the assumption that choice is transitive conditional on history H. I could instead have assumed that the choice function conditions on an object individuated at a less fine level. For example it could condition just on the multiset of all attributes encountered along each dimension. There are many variations between these two extremes, and the correct structure of the object H is an empirical question: e.g., does encountering a \$2 price have the same effect on preferences no matter in what context it was observed?

There are two conditions on H which are of particular interest. First, does choice depend on the order in which choice sets are encountered? (If not, then H could be defined as a multiset instead of a vector). If the composition of choice sets is treated as conveying information (as in Wernerfelt (1995), Kamenica (2008)), and choice sets are independent draws from a distribution of choice sets, then the order should not matter. However a number of studies do find order effects, for example Bhargava (2008), as discussed, finds that more recent decisions have a bigger influence. On the other hand Levin et al. (1987) and Ariely et al. (2003) seem to find that earlier choice sets have more influence. This distinction, whether order matters, seems a natural way of distinguishing between behaviour that seems to be "preference discovery" (Plott (1996)), and "preference construction" (Slovic (1995), Ariely et al. (2003)).

Second, of independent interest is whether preferences converge to being transitive. If, for example, we defined H as either the set of unique choice sets encountered, or the ordered set of unique choice sets, then exposing subjects to a repeated sequence of choice sets would lead to, eventually, decisions which are sequentially transitive as well as jointly transitive. Plott (1996) surveys a range of experimental evidence to argue that preferences tend to converge to the rational benchmark with experience.

5.2 Models of Menu Effects

The model presented here is closely related to models of menu effects, where the utility function depends on the choice set, i.e. models which have a utility function like:

$$U(x, A)$$

An early example was Tversky and Simonson (1993). These models have been popular in the marketing literature, used to explain decoy effects, such as, for example Mellers and Cooke (1994). Two notable recent papers are Bordalo et al. (2011), and Kőszegi and Szeidl (2011). Also the

Kőszegi and Rabin (2006) model of endogenous reference points can be considered a member of this class of models.

The model given here differs from the models just mentioned in a number of ways. First, all of the above models assume a separable utility function:

$$U(x, A) = \sum_i f_i(x_i, A)$$

and some assume a linear (Kőszegi and Szeidl (2011)) or piecewise linear (Kőszegi and Rabin (2006), Bordalo et al. (2011)) form for f_i. The model presented in this paper makes much weaker assumptions about the shape of the utility function. This allows the comparison effect to be combined with other restrictions on the utility function. For example, consider the following common functional form assumptions (representing exponential discounting, expected utility, and quasi-linearity in money, respectively):

$$\begin{aligned} U(x,t) &= e^{-rt}u(x) \\ U(p,x) &= p \cdot u(x) \\ U(x,m) &= u(x) - m \end{aligned}$$

In each case the model can be made consistent with a comparison effect, by modifying it in the following way:

$$\begin{aligned} U(x,t) &= e^{-\gamma(H)rt}u(x) \\ U(p,x) &= p \cdot \gamma(H) \cdot u(x) \\ U(x,m) &= u(x) - \gamma(H) \cdot m \end{aligned}$$

Where $\gamma(H)$ is a function of the choice history.

A second advantage of the model presented in this paper is that for almost all of the predictions derived above, it is not necessary to know the full choice set.[20] The comparison effect assumes that each element in H has a monotonic effect on the marginal rate of substitution, independent of the rest of H. This is not true for any of the other theories, for example the predictions of Bordalo et al. (2011) are contingent on knowing the average value along each dimension, and Kőszegi and Szeidl (2011) require knowledge of the maximum and minimum along each dimension.

Third, unlike the other models I have explicitly taken into account the effect of comparisons outside the choice set, by conditioning on history. Most of the papers above do cite evidence from variation outside the choice set, but this is done in an *ad hoc* way. More important than simply formalising the choice history, H, is my assumption that decision-makers use a consistent utility function when making decisions from joint choices. If this is accepted (and I have noted that the

[20] The exception is the joint-separate difference in levels of WTP. There I assume that the marginal rate of substitution depends on the average, thus to predict that the MRS is affected in a particular direction requires knowing the previous average.

evidence seems to support it), then joint choice offers a powerful channel for testing models with menu effects: by manipulating choice set $A_{t,1}$, and observing choice from $A_{t,2}$, menu effects can be easily identified.

An important related paper, though less formal, is Hsee and Zhang (2010), which outlines a theory of "evaluability".[21] They say that subjects will exhibit insensitivity to changes in a dimension if the following three conditions hold (i) the evaluation mode is separate, rather than joint; (ii) subjects know relatively little about the domain; (iii) the domain is not intrinsically evaluable (people do not have an "innate and stable physiological or psychological scale (reference system) to evaluate values on an attribute"). They discuss three results: that scope insensitivity will occur in separate but not joint evaluation; that ranking reversals can occur between joint and separate judgments of WTP, if one dimension is less evaluable; and that subjects tend to mispredict future preferences, depending on the evaluation mode at the time of prediction. The first two results are obviously closely related to results derived in this paper. The model of comparison effects presented here can thus be thought of as, in part, a formalisation of Hsee and Zhang's "evaluability theory".

5.3 Reference Point and Position of the Origin

The predictions of the model presented in this paper are sensitive to the choice of a reference point, and to the choice of dimensions of comparison. However the sensitivity is not severe.

Choices are often described in terms of *changes* to your outcome relative to your initial state, whereas in standard decision theory preferences are defined over final outcomes, i.e. your initial state plus the change. This can create an ambiguity in the model presented here if some alternatives contain negative changes. For example when making a decision about a purchase we generally describe alternatives with the price of each good, though we could equivalently describe instead the terminal wealth (i.e., initial wealth minus price). This would change the predictions of the model: for example increasing the price of some alternative would, under the first description, lower sensitivity to prices; but in the second description, because it would lower terminal wealth, it would lower sensitivity to money. Note that this is only important for translations in the reference point which change the sign of attributes; if the sign is preserved, then likewise the sign of any comparative static prediction is preserved.

This is one reason why I have largely avoided discussing experiments relating to probabilities, because a prospect of a gain with probability p can also be thought of as a loss with probability $1 - p$, thus any change in the sensitivity to probability will have different effects depending on how the gamble is conceptualized by the subject. Nevertheless, the general patterns in choice between lotteries seem to fit: experiments using matching and using choice with a common difference in probability (i.e., Allais type questions) both seem to find diminishing sensitivity in probability, as suggested by the model here (for data see, for example, (Bordalo et al., 2010)).

A second limitation is that the predictions are sensitive to a choice of dimensions. For example a point on a plane can be defined either using a Cartesian or a polar coordinate system. Our definition

[21] Building on earlier work on joint and separate evaluation, Hsee et al. (1999).

of relative thinking preferences implicitly assumes a canonical set of axes.

One final complication in interpretation is the behaviour of comparisons in the extreme. This is a difficulty for almost all theories of comparison: it seems sensible to assume that if a comparator becomes infinitely large or small, then it is ignored (e.g. increasing a comparison from \$10 to \$1 trillion might affect choice less than increasing from \$10 to \$100). But then if $\lim_{x\to-\infty} f(x) = \lim_{x\to+\infty} f(x)$, and $f'(x) \neq 0$, then f must be non-monotonic. In other words, if the comparator effect diminishes in the extremes, then it must be non-monotonic somewhere inside the extremes. In fact the few experiments performed have found that even for quite large magnitudes observed effects do not reverse, or only very slightly, see e.g. Krishna et al. (2006) and Chapman and Johnson (1994).

6 Comparisons and Market Equilibrium

The framework and the comparison effect can be naturally embedded in a standard market equilibrium model.

For a monopolist the implications are straight-forward: a consumer's willingness to pay (and their demand) can be increased by manipulating comparisons, either by raising the price of the comparison (making your product seem cheap by the contrast), or by lowering the quality or quantity of the comparison (making your product seem to have a high quality by the contrast).

In the rest of this section I embed the theory in a model of oligopoly with unit demand.

6.1 Unit Demand Equilibrium

I begin with a baseline model to show how there will be perfect pass-through of costs in a model of differentiation under two conditions used in the standard Hotelling model: (i) unit demand (meaning that every consumer buys exactly one product, i.e. total demand is inelastic); and (ii) money is linearly separable in the utility function, i.e. there are no income effects. All proofs are in an Appendix.

Suppose there are N firms, and a unit mass of consumers, indexed with j. Each consumer chooses to purchase one firm's product, from the vector of prices $p = (p_1, ..., p_N) \in \mathbb{R}^N$, and each consumer has a vector of idiosyncratic valuations for each of the N products, $v_j = (v_{j,1}, ..., v_{j,N}) \in \mathbb{R}^N$. The valuations are drawn from the continuously differentiable joint distribution F.

I assume quasilinear utility, i.e. the utility that consumer j receives from purchasing product i is assumed to be:[22]

$$U_{j,i} = v_{j,i} + \beta(m_j - p_i)$$

where m_j is the consumer's wealth, and $v_{j,i}$ is consumer j's idiosyncratic value for good i. For now I consider β to be a constant; in the next section I will allow β to depend on the prices observed.[23]

[22] This is an indirect utility function, equivalent to a utility function which is linear in consumption of the outside good.

[23] The assumption that consumers share a common β is without loss of generality, because of the lack of restriction on the distribution of v.

Each consumer will thus choose the good i which maximises the term $v_{j,i} - \beta p_i$. I assume that in equilibrium the maximum value of $v_{j,i} - \beta p_i$ is always greater than an outside option $\underline{u} = 0$; this assumption makes total demand inelastic - i.e., every customer buys exactly one good.

Demand can be written as an integral over the density of preferences:

$$D^i(p) = \int_{\beta p_i}^{\infty} \int_{\prod_{j \neq i} S_{j,i}} f(v) dv_{-i} dv_i$$

Where $S_{j,i} = (-\infty, \beta p_j + v_i - \beta p_i)$, which represents the set of valuations of good j such that, at given prices, good i will be preferred to good j. By our assumption of unit demand, $v_{j,i} > \beta p_i$, so we can simplify this to

$$D^i(p) = \int_{-\infty}^{\infty} \int_{\prod_{j \neq i} S_{j,i}} f(v) dv_{-i} dv_i$$

This says that the total density inside the intersection of $S_{j,i}$ for all $j \neq i$ is equal to the demand for good i. Note that this demand function is invariant to a constant addition to all prices, i.e. it is preserved when prices are transformed $p_i = p_i + \tau$. In words, consumers' decisions are determined only by the differences between prices.

Turn now to the producers. Each producer faces the same marginal cost, c, and all choose their price simultaneously, prior to consumers' demand decisions. They thus have a profit function

$$\pi_i = (p_i - c) D^i(p)$$

With first-order condition,

$$(p_i - c) D^{ii}(p) + D^i(p) = 0$$

where $D^{ii} = \frac{\partial}{\partial p_i} D^i(p)$.

The model so far allows great flexibility in the demand system, and does not guarantee either existence or uniqueness of an equilibrium set of prices. For simplicity I therefore assume that demand is twice differentiable, and the Jacobian of the demand system is negative definite, which guarantees existence and uniqueness (Vives (2001), p145).[24] The model thus encompasses both horizontal and vertical differentiation.[25]

This allows our first comparative static,

Proposition 8. *Without relative thinking* $\frac{dp_i}{dc} = 1$ *for all* i.

This simply states that, because demand depends only on *differences* between prices, the equilib-

[24] I will assume this holds both for the ordinary demand function and the demand function with relative thinking. Caplin and Nalebuff (1991)have an extended discussion of conditions on utility functions which will underlie a demand system with an equilibrium.

[25] Typically in vertical differentiation models consumers have the same preferred ranking among alternatives; in horizontal differentiation, they have different rankings. Because I put so few constraints on F, either can occur in this model. In a two-firm Hotelling model with uniform customers and quadratic transport costs (Tirole (1994) p281), the difference in utilities $u_{j,1} - u_{j,2}$ would be distributed uniformly.

rium in relative prices will be independent of the level of cost, in other words aggregate pass-through will be exactly 1.

6.2 Demand with a Comparison Effect

I now introduce a utility function with choice-set dependence. In particular, it is quasilinear in money for a *given* choice set, but sensitivity to money decreases with the magnitude of elements in the choice set.[26]

$$U_{j,i} = v_{j,i} + \beta(p_1, ..., p_N)\,[m_j - p_i]$$

with

$$\beta > 0,\ \frac{\partial \beta(p_1, ..., p_N)}{\partial p_k} < 0,\ \forall k$$

The marginal rate of substitution between goods and money is now

$$MRS_{v,p} = \frac{\partial U/\partial p}{\partial U/\partial v} = -\beta(p_1, ..., p_N)$$

Thus, because β is decreasing in every price, the utility function exhibits the comparison effect.

For now I will assume that firms do not internalise their own effect on β; relaxing this assumption is discussed later. Though β may vary between customers, our results here do depend on $\frac{\partial \beta}{\partial p_i}$ being the same for all customers. Similar results could be derived if I allowed $\frac{\partial \beta}{\partial p_i}$ to vary between customers, but we would require restrictions on the distribution of v, and its covariance with $\frac{\partial \beta}{\partial p_i}$.

Using this representation, pass-through will be greater than one, i.e., higher cost goods will have higher markups:

Proposition 9. *With relative thinking then* $\frac{dp_i}{dc} \geq 1$*, for all* i*, i.e. aggregate pass-through is greater than one.*

This holds simply because when prices are higher, customers become less sensitive to price differences (in the Hotelling model, it is equivalent to an increase in transport costs). For every firm, the sensitivity of demand to price becomes less, so they raise their price.

Proposition 10. *With relative thinking then for any* i *and* j*,* $\frac{|dp_i - dp_j|}{dc} \geq 0$*, i.e. dispersion is increasing in cost.*

The result may be thought to be surprisingly unambiguous, given the lack of assumptions on demand. Mathematically it holds because demand depends only on weighted price differences $\beta(p_j - p_i)$. Thus, if all prices increase by the same fixed amount, the first-order conditions will still hold. Likewise if β increases by some fixed factor, and all price differences decrease by the same factor (equivalently, if all margins decrease by this factor) then the terms $\beta(p_j - p_i)$ will remain unchanged, and the first-order conditions will hold.

[26] See Cunningham (2011) for an explanation of how joint choice experiments can identify choice-set effects. In this case quasilinear preferences can be identified if subjects, when presented with two choice sets jointly, make no preference reversals when one choice set differs from the other only in a common difference in prices.

Intuitively, when consumers become 10% less sensitive to prices (β falls by 10%), then having all firms increase their margins by 10% will return demand and marginal demand to exactly the same position as before, restoring equilibrium.

6.3 Endogenous Entry

So far I have assumed a fixed number of firms, implying that the profit earned will vary with cost (because cost increases markups, without affecting quantity sold). Free entry will naturally eliminate the effect of cost on profits. Nevertheless, as I will show, the effect on markups remains, at least in the simplified case of symmetric competition on a circle (a Salop model).

For simplicity I assume a large market and thus treat the number of firms n as a continuous variable.

Proposition 11. *When customers are distributed uniformly on a circle, with quadratic transport costs, and the number of firms n is determined by a fixed cost C and a zero profit condition, then*

$$\begin{aligned} \frac{dn}{dc} &> 0 \\ \frac{dp}{dc} &> 1 \end{aligned}$$

The lower sensitivity to price is now taken up in two ways: firms charge higher markups, and more firms enter.

Thus for high-priced goods which are sold at multiple outlets, we should see many retailers, each selling relatively few goods. This may be true in some markets, e.g. jewellers, car dealers, estate agents, optometrists, where it could be thought that there are a surprising number of low-volume, high-margin outlets. As with most cross-industry predictions, this is an extremely difficult proposition to test, because many other characteristics important for industry structure are likely to covary with the cost of the good being sold (Sutton (1992)).

6.4 Endogenising β

The model presented does not allow firms to take into account the effect of their own price on price-sensitivity, β. I will discuss briefly the consequences of this in a 2-firm model.

Call the two firms H and L, with names assigned such that $p_H \geq p_L$. Let the proportion of customers who buy good L be given by $F(\beta(p_H - p_L))$. The two profit functions and first-order conditions will be

$$\begin{aligned} \pi_L &= (p_L - c)F\left(\beta(p_H - p_L)\right) \\ \pi_H &= (p_H - c)\left[1 - F\left(\beta(p_H - p_L)\right)\right] \\ \pi_L' &= F - (p_L - c)\beta F' + \frac{\partial \beta}{\partial p_L}(p_H - p_L)(p_L - c)F \\ \pi_H' &= 1 - F - (p_H - c)\beta F' - \frac{\partial \beta}{\partial p_H}(p_H - p_L)(p_H - c)F' \end{aligned}$$

Only the final term in each first-order condition is new. Because $\frac{\partial \beta}{\partial p_L} < 0$, the extra term must be negative for firm L, indicating an incentive to lower their price. A lower price makes every consumer more sensitive to a given difference in prices, and the marginal consumer therefore switches to the low-price firm. The corresponding term is positive for firm H because they are better off when the marginal consumer becomes *less* sensitive to a given difference in prices. In a symmetric equilibrium where $p_H = p_L$ the term disappears because neither firm cares about their influence on β; a change in β will not affect the choice of any customer.

In this case endogenising β gives an extra incentive towards price dispersion. This is of relevance if $\beta(\bar{p})$ is decreasing and convex, because as c rises, and prices rise, then β' will become smaller, lowering the incentives for dispersion. The net effect may then go in the opposite direction from Proposition 3. It remains to be shown under what conditions the direct effect dominates this indirect effect.

6.5 Income Effects

I have assumed that the underlying utility function is linearly separable in money for a given β, i.e. the consumer is risk neutral. Introducing concave utility for money is not trivial, but it is likely to reinforce the results. As prices increase, terminal wealth decreases, so the marginal utility of money increases. The effect thus is similar to an increase in β, raising the sensitivity to differences in price, and thus lowering equilibrium markups and equilibrium dispersion - i.e. the effect is in the opposite direction from that predicted by relative thinking. Put another way, risk aversion does not seem to help explain the observed positive relationship between price and markup, if anything it would seem to predict the opposite.

7 Evidence

Here I survey the evidence for the first two predictions: that markup is increasing in cost, and that dispersion is increasing in cost.

To summarise, there is strong evidence for both effects in the cross-section of products. However in cross-section studies identification is not strong, I discuss a variety of possible confounding factors. In time series studies identification is much stronger, and published studies largely supports the model's predictions for markup. Unfortunately I am not aware of any time-series studies which look at price dispersion.

The paper's predictions for prices are driven through relative thinking's distortions of the set of demand functions. We are thus indirectly looking for features of demand by observing prices. There are then two factors we should consider in interpreting the evidence: first, whether the inferred shape of demand could occur without relative thinking. Second, whether observed prices may not reveal demand, due to other confounding factors.

Tackling the first problem, the prediction can be stated as pass-through rates being greater than one ($\frac{dp}{dc} > 1$). Pass-through rates depend on the industry structure. If the industry is perfectly competitive then the rate of pass through must be less than or equal to 1 (as long as demand slopes down and supply slopes up). At the other extreme, for a monopolist, their pass-through rate depends on the curvature of the demand curve they face: it will be greater than 1 if and only if the demand curve is log convex.[27] Of interest here is the oligopoly case, because all the studies I cite consider homogenous goods sold at different outlets. As I have shown in the previous sections, with unit demand and quasilinear preferences, then pass-through will always be equal to 1. Any deviation must therefore be due to a violation of unit demand, i.e. total demand varying with cost.[28] In a symmetric model this must mean that as prices rise all firms face a lower ratio of marginal consumers than inframarginal, i.e. demand has a decreasing hazard rate (a.k.a. a heavy-tailed distribution). I cannot rule this out as an alternative explanation of the results.

Turning to price dispersion, most models predict that dispersion is independent of cost, because most assume unit demand and quasilinear preferences: i.e., they set up the problem as choosing which store to buy an item from, ignoring the question of how many items to buy. The problem is thus entirely independent of cost, and this holds for dispersion driven by differentiation (as in this paper), or dispersion driven by imperfect information (see Baye et al. (2006) for a survey). It is not clear what can be said about cost and dispersion when the assumption of unit demand is relaxed.

7.1 Evidence on Markups

Because cost data is difficult to obtain (and because of endogeneity problems) estimates of pass-through have often been based on tax changes, which largely find pass-through rates greater than one. For cigarette taxes Barzel (1976) found pass-through slightly greater than one. For alcohol taxes Kenkel (2005) and Young and Bielinska-Kwapisz (2002) both find pass-through greater than one. Estimating pass-through using changes in broad-based sales taxes, Poterba (1996) finds pass-through close to 1, but Besley and Rosen (1999) find a higher pass-through.

7.2 Evidence from a Drugstore

I have collected a new dataset, which documents the cross-section relationship of cost and markup. The data are 3,500 observations of cost and price from a Cambridge, Massachusetts branch of a national drugstore chain. It is unusual to observe cost for a retailer: a comparable dataset is used in Eichenbaum et al. (2011), who say "we're really, I think, one of the first people to ever get data on marginal costs and you see fascinating patterns" (Eichenbaum and Vaitilingam (2009)). However Eichenbaum et al. are not able to publish data on markups, as I am, because they say "[o]ur agreement with the retailer does not permit us to report information about the level of the markup for any one item or group of items."

Our dataset is documented further in an appendix. Figure 3 plots item cost against item markup, and shows the very strong relationship between the two variables. On average \$1.00 increase in cost

[27] The monopolist's first order condition is $(p-c) = -Q(p)/Q'(p)$, so $\frac{dp}{dc} = \frac{1}{1+\frac{\partial}{\partial p}\frac{Q(p)}{Q'(p)}}$, and $\frac{\partial^2}{\partial^2 p} log(Q(p)) = \frac{\partial}{\partial p}\frac{Q'(p)}{Q(p)}$, thus log convexity or concavity determines the pass-through rate.

[28] I discussed earlier why violations of quasi-linearity are likely to push in the other direction.

is associated with an increase in absolute markup of \$0.73, thus proportional markup is decreasing in cost, although slowly.

One feature in particular is notable: the absence of outliers, i.e. goods with either low-cost and high-markup, or high-cost and low-markup. If the cost-markup relationship was driven by a correlation between cost and demographics, the correlation must be extremely strong. Put another way, the graph would imply that there are almost no low-cost goods with high-demographic customers, or high-cost goods with low-demographic customers. As examples, various types of branded lip balm, coffee filters, and scented candles (plausibly high-demographic goods) all have cost below \$1 and markup below \$1. Whereas only two products out of 400 with cost above \$10 have a markup below \$1.[29]

Because this is cross-section evidence, the cost-markup relationship does not directly identify the shape of the demand functions: it could be driven by confounding factors. Three factors are worth discussing: the demographics of the customers, the substitutability with other goods sold by the firm, and the frequency of purchase. All three factors are reasons why demand might be less sensitive for high cost items, so they could predict both the markup and dispersion relationships.

First, if high-cost goods were bought by customers with less sensitive demand (lower hazard rates), this would cause them to be associated with a higher markup. An important determinant of demand sensitivity should be income, determining the opportunity cost of your time, thus if high-cost items are bought by high-income customers, I expect the positive predicted relationship.

Second, if firms sell multiple products then the optimal markup depends on interactions: if low-cost goods tend to be more complementary with other goods than are high-cost goods, this would again predict the observed relationship. A strong source of complementarity would come from any fixed cost of a visit to a store because, in effect, if a good is often purchased in a basket with other goods, the total markup can be spread out over the total basket. Thus if low-cost goods are more often bought in a basket with other goods, this will also generate a positive cost-markup relationship in the cross-section.

Third, if a good is more frequently purchased, there is a stronger incentive to collect price information, leading to a higher price sensitivity (Sorensen (2000)). If high-cost goods tend to be purchased less frequently, high cost goods should thus have higher markups, and higher price dispersion.

Finally, there are some biases which may affect estimation. First, the observed marginal cost of goods does not account for other variable costs, such as handling. If high cost goods tend to have higher handling costs, this would also account for higher *measured* margins, even when the true margins are the same. (One handling cost certainly higher for high-cost goods is the cost of capital). Second, there may be some censoring due to turnover. If a good is sold infrequently, the opportunity cost of allocating shelf-space to this good may be too high. Thus if there is a negative correlation between cost and turnover, we should expect to observe positive correlation between cost and markup in products observed due to censoring. Third, firms may simply use a rule of thumb in

[29] Huggies nappies and Pampers nappies.

pricing, marking up by a constant fraction, even when it departs from the profit maximising price. Looking again at figure 3, if firms used a rule of thumb we would expect observations to be lined up along upward sloping lines. Instead we see much variation for products with the same cost, indicating that the firm conditions on information other than cost when setting price.

7.3 Evidence on Dispersion

There are a number of empirical papers which document the correlates of price dispersion. Most papers do not have access to cost data, but they do report the relationship between price and dispersion. Unfortunately I do not know of any papers about dispersion which use a strongly exogenous source of cost variation, as in the literature above which measures the effect of tax changes on markups. Instead all the papers I describe here are in the cross section, and so are subject to the caveats that I have already mentioned.

In short, every study I know of has found a very strong positive relationship between price and price dispersion. The effect is so strong that a number of papers use *proportional* dispersion as a measure, i.e. using $\frac{p_1-p_2}{\frac{1}{2}(p_1+p_2)}$ instead of $(p_1 - p_2)$. If proportional price dispersion is constant in cost, then absolute dispersion is increasing in cost. However to be consistent with the models they cite, these papers should be measuring absolute price dispersion (see, for example, Baye et al. (2004), Clay et al. (2002), Jaeger and Storchmann (2011), Lach (2002)). Confusion on this point leads to some illogical statements.[30]

An early paper on price dispersion was Pratt et al. (1979), using a variety of different goods, and they find a strong positive relationship between price and dispersion. So do Aalto-Setala (2003) for groceries, Baye et al. (2004) and Pan et al. (2001) for goods sold online, Clay et al. (2002) for books sold online, Hoomissen (1988) and Lach (2002) for consumer goods, Jaeger and Storchmann (2011) for wine, and Sorensen (2000) for prescription drugs. The empirical literature is surveyed in Ratchford (2009).

The same caveats apply from last section, regarding correlates of cost. However the two studies which control for purchase frequency finds that it makes very little change to the estimated relationship between dispersion and price. In Pratt et al. (1979), the estimated coefficient of $ln(price)$ on $s.d.(price)$ shrank from 0.892 to 0.836, when a crude control for purchase frequency was introduced. Sorensen (2000) reports results using both cost and purchase frequency as explanatory variables, and finds that a $1 higher cost is associated with a 20 cent increase in the range of prices.

7.4 Related Literature

This IO application of the comparative thinking model paper broadly belongs to a family of recent literature examining the effects of non-standard decision making in different market equilibrium settings, surveyed in Ellison (2006) and Spiegler (2011).

[30] For example, Clay et al. (2002) say "The increase in standard deviation with price is ... somewhat surprising ... given that search models predict that customers will engage in more search for higher priced items and so price dispersion will be lower." Lach (2002) says "search costs are low relative to the high price of the good and, as a consequence, more searching for the lowest price is undertaken."

The idea that proportional thinking may help explain patterns in price dispersion has been brought up a number of times, first by Tversky and Kahneman (1981), in the course of introducing the jacket/calculator example, who mention that it may explain the relationship between price and price dispersion found in Pratt et al. (1979). Grewal and Marmorstein (1994) make the same claim, and report that willingness to search seems related to the base price of a good, and that dispersion seems to be increasing the price of goods.

Most closely related to this paper is Azar (2008), which constructs a 2-firm model with horizontal differentiation. Azar shows that in that model dispersion is increasing in transport costs, and proposes that relative thinking can be modeled as transport costs increasing in price.[31] The model presented here extends Azar's work in a number of ways: in deriving the behavioural results from a general model of relative thinking, in using a much more general model, in deriving predictions for both the level and dispersion of prices, as well as firm entry, and finally in introducing new data on the level of markups.

8 Conclusion

This paper has been based around a simple framework for describing how choices are affected by irrelevant information, through defining a choice function $c(A, H)$, which allows choice to depend on the history of choice sets. Many patterns of anomalous choice can be described in this framework. I further assumed that, conditional on history, choices can be represented with a utility function, an assumption that seems to be supported by the evidence.

Within this framework I then showed that a common pattern seems to fit very many of the anomalous phenomena: that an increase in an attribute in the history tends to lower sensitivity to marginal differences along that dimension.

The postulated comparison effect is similar to effects found in perceptual judgment. Recently cognitive scientists have argued that these perceptual effects may reflect optimal inference given a limited information set, and I give sufficient conditions under which the contrast effect will emerge from Bayesian inference.

The second half of the paper has shown how comparison effects will change a system of demand functions if the choice set consists of a product available at different stores. The new demand system, which would ordinarily imply constant markups, now implies that in equilibrium higher cost will be associated with higher markups, higher dispersion, and more entry. I have shown that data supports the predictions, both in time series and in cross-section, though other explanations are difficult to rule out.

Finally I introduced a large dataset of costs and prices from a drugstore, which shows a very tight relationship between cost and price, as predicted by the model.

[31] Using this interpretation, Azar (2008) says that higher prices can have negative welfare effects through raising the unpleasantness of travel. It seems more natural to assume, as in this paper, that relative thinking works through higher prices lowering the subjective value of *money*, rather than raising the subjective value of transport. The distinction is useful for welfare analysis, but also if we are to predict how tradeoffs are made against other goods apart from time or money.

References

Aalto-Setala, V. (2003), "Explaining price dispersion for homogeneous grocery products." *Journal of Agricultural & Food Industrial Organization*, 1, 9.

Allais, M. (1953), "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine." *Econometrica*, 31, 503–546.

Ariely, D., G. Loewenstein, and D. Prelec (2003), "Coherent arbitrariness: Stable demand curves without stable preferences." *Quarterly Journal of Economics*, 118, 73–105.

Azar, O.H. (2008), "The effect of relative thinking on firm strategy and market outcomes: A location differentiation model with endogenous transportation costs." *Journal of Economic Psychology*, 29, 684–697.

Azar, O.H. (2009), "Does relative thinking exist in real-world situations? a field experiment with bagels and cream cheese." *Economic Inquiry*, 49, 564–572.

Azar, O.H. (2011), "Do consumers make too much effort to save on cheap items and too little to save on expensive items? experimental results and implications for business strategy." *American Behavioral Scientist*, 55, 1077–1098.

Bartels, D.M. (2006), "Proportion dominance: The generality and variability of favoring relative savings over absolute savings." *Organizational Behavior and Human Decision Processes*, 100, 76–95.

Barzel, Y. (1976), "An alternative approach to the analysis of taxation." *The Journal of Political Economy*, 84, 1177.

Baye, M.R., J. Morgan, and P. Scholten (2004), "Price dispersion in the small and in the large: Evidence from an internet price comparison site." *The Journal of Industrial Economics*, 52, 463–496.

Baye, M.R., J. Morgan, and P. Scholten (2006), "Information, search, and price dispersion." In *Handbook of Economics and Information Systems* (T. Hendershott, ed.), Elsevier Press, Amsterdam.

Beauchamp, J.P., D.J. Benjamin, C.F. Chabris, and D.I. Laibson (2011), "How malleable are risk preferences and loss aversion?"

Becker, G.M., M.H. DeGroot, and J. Marschak (1964), "Measuring utility by a single-response sequential method." *Behavioral Science*, 9, 226–232.

Besley, T. and H. Rosen (1999), "Sales taxes and prices: an empirical analysis." *National Tax Journal*, 52, 157–178.

Bhargava, S. (2008), "Perception is relative: Sequential contrasts in the field." UC Berkeley Working Paper.

Birnbaum, M.H. (1992), "Violations of monotonicity and contextual effects in choice-based certainty equivalents." *Psychological Science*, 3, 310.

Birnbaum, M.H., J.N. Patton, and M.K. Lott (1999), "Evidence against rank-dependent utility theories: Tests of cumulative independence, interval independence, stochastic dominance, and transitivity." *Organizational Behavior and Human Decision Processes*, 77, 44–83.

Bordalo, P., N. Gennaioli, and A. Shleifer (2010), "Salience theory of choice under risk."

Bordalo, P., N. Gennaioli, and A. Shleifer (2011), "Salience and consumer choice." Working Paper.

Caplin, A. and B. Nalebuff (1991), "Aggregation and imperfect competition: On the existence of equilibrium." *Econometrica*, 25–59.

Carson, R.T. (1997), "Contingent valuation surveys and tests of insensitivity to scope." In *Determining the value of non-marketed goods: Economic, psychological, and policy relevant aspects of contingent valuation methods* (R.J. Kopp, W. Pommerhene, and N. Schwartz, eds.), 63, Kluwer Academic Publishers.

Chapman, G.B. and E.J. Johnson (1994), "The limits of anchoring." *Journal of Behavioral Decision Making*, 7, 223–242.

Chapman, G.B. and E.J. Johnson (2002), "Incorporating the irrelevant: Anchors in judgments of belief and value." In *Heuristics and biases: The psychology of intuitive judgment* (T. Gilovich, D. Griffin, and D. Kahneman, eds.).

Clay, K., R. Krishnan, E. Wolff, and D. Fernandes (2002), "Retail strategies on the web: Price and non–price competition in the online book industry." *The Journal of Industrial Economics*, 50, 351–367.

Cunningham, T.E. (2011), "Joint choice and relative judgment." Dissertation Chapter, London School of Economics.

Day, B. and J.L. Pinto Prades (2010), "Ordering anomalies in choice experiments." *Journal of Environmental Economics and Management*, 59, 271–285.

Desvousges, W.H., F.R. Johnson, R.W. Dunford, K.J. Boyle, S.P. Hudson, and K.N. Wilson (1993), "Measuring natural resource damages with contingent valuation: tests of validity and reliability." In *Contingent valuation: A critical assessment* (J. A. Hausman, ed.), 91–159, Amsterdam: North Holland.

Diamond, P.A. and J.A. Hausman (1994), "Contingent valuation: is some number better than no number?" *The Journal of Economic Perspectives*, 8, 45–64.

Eichenbaum, M., N. Jaimovich, and S. Rebelo (2011), "Reference prices, costs, and nominal rigidities." *American Economic Review*, 101, 234–262.

Eichenbaum, M. and R. Vaitilingam (2009), "Nominal rigidities: how often do retailers really change prices?" Vox EU, Article 3611.

Ellison, G. (2006), "Bounded rationality in industrial organization." In *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, volume 2, 142–74, Cambridge, UK: Cambridge University Press.

Frederick, S. and B. Fischhoff (1998), "Scope (in) sensitivity in elicited valuations." *Risk Decision and Policy*, 3, 109–123.

Frederick, S., G. Loewenstein, and T. O'donoghue (2002), "Time discounting and time preference: A critical review." *Journal of economic literature*, 40, 351–401.

Frederick, S. W. and D. Mochon (2011), "A scale distortion theory of anchoring." *Journal of Experimental Psychology: General*.

Fudenberg, D., D.K. Levine, and Z. Maniadis (2010), "Reexamining coherent arbitrariness for the evaluation of common goods and simple lotteries."

Green, L., N. Fristoe, and J. Myerson (1994), "Temporal discounting and preference reversals in choice between delayed outcomes." *Psychonomic Bulletin & Review*, 1, 383–389.

Grewal, D. and H. Marmorstein (1994), "Market price variation, perceived price variation, and consumers' price search decisions for durable goods." *Journal of Consumer Research*, 21, 453–460.

Herne, K. (1999), "The effects of decoy gambles on individual choice." *Experimental Economics*, 2, 31–40.

Hoomissen, Theresa Van (1988), "Price dispersion and inflation: Evidence from israel." *Journal of Political Economy*, 96, pp. 1303–1314.

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

Hsee, C.K., G.F. Loewenstein, S. Blount, and M.H. Bazerman (1999), "Preference reversals between joint and separate evaluations of options: A review and theoretical analysis." *Psychological Bulletin*, 125, 576.

Hsee, C.K. and J. Zhang (2010), "General evaluability theory." *Perspectives on Psychological Science*, 5, 343.

Huber, J., J.W. Payne, and C. Puto (1982), "Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis." *Journal of Consumer Research*, 9, 90–98.

Jaeger, D.A. and K. Storchmann (2011), "Wine retail price dispersion in the united states: Searching for expensive wines?" *American Economic Review*, 101, 136–41.

Jahedi, S. (2008), "A taste for bargains." Unpublished Working Paper.

James, W. (1890), *The Principles of Psychology, Vol I.* Henry Holt and Co.

Kahneman, D. and S. Frederick (2002), "Representativeness revisited: Attribute substitution in intuitive judgment." In *Heuristics and biases: The psychology of intuitive judgment*, 49–81.

Kahneman, D. and S. Frederick (2005), "A model of heuristic judgment." *The Cambridge handbook of thinking and reasoning*, 267–294.

Kahneman, D., I. Ritov, and D. Schkade (1999), "Economic preferences or attitude expressions?: An analysis of dollar responses to public issues." *Journal of Risk and Uncertainty*, 19, 203–235.

Kamenica, E. (2008), "Contextual inference in markets: On the informational content of product lines." *American Economic Review*, 98, 2127–2149.

Kenkel, D.S. (2005), "Are alcohol tax hikes fully passed through to prices? evidence from alaska." *The American economic review*, 95, 273–277.

Kersten, D., P. Mamassian, and A. Yuille (2004), "Object perception as bayesian inference." *Annu. Rev. Psychol.*, 55, 271–304.

Kőszegi, B. and M. Rabin (2006), "A model of reference-dependent preferences." *The Quarterly Journal of Economics*, 121, 1133–1165.

Kőszegi, B. and A. Szeidl (2011), "A model of focusing in economic choice." Working Paper.

Krishna, A., M. Wagner, C. Yoon, and R. Adaval (2006), "Effects of extreme-priced products on consumer reservation prices." *Journal of Consumer Psychology*, 16, 179–193.

Lach, S. (2002), "Existence and persistence of price dispersion: an empirical analysis." *Review of Economics and Statistics*, 84, 433–444.

Levin, I.P., R.D. Johnson, and M.L. Davis (1987), "How information frame influences risky decisions: Between-subjects and within-subject comparisons* 1." *Journal of economic psychology*, 8, 43–54.

Lichtenstein, S. and P. Slovic (1971), "Reversals of preference between bids and choices in gambling decisions." *Journal of experimental psychology*, 89, 46.

Mas-Colell, A., M.D. Whinston, and J.R. Green (1995), *Microeconomic theory.* Oxford University Press, New York.

Mazar, N., B. Kőszegi, and D. Ariely (2009), "Price-sensitive preferences." *Working Paper.*

Mazumdar, T., SP Raj, and I. Sinha (2005), “Reference price research: Review and propositions.” *Journal of Marketing*, 69, 84–102.

Mellers, B.A. and A.D.J. Cooke (1994), “Trade-offs depend on attribute range.” *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1055.

Milgrom, P.R. (1981), “Good news and bad news: Representation theorems and applications.” *The Bell Journal of Economics*, 12, 380–391.

Mochon, D. and S. Frederick (2011), “Anchoring in sequential judgments.” *Working Paper*.

Ordóñez, L.D., B.A. Mellers, S.J. Chang, and J. Roberts (1995), “Are preference reversals reduced when made explicit?” *Journal of Behavioral Decision Making*, 8, 265–277.

Pan, X., B.T. Ratchford, and V. Shankar (2001), “Why aren’t the prices of the same item the same at me.com and you. com?: Drivers of price dispersion among e-tailers.” Working Paper.

Parducci, A. (1995), *Happiness, pleasure, and judgment: The contextual theory and its applications*. Lawrence Erlbaum, Mahwah, NJ.

Plott, C.R. (1996), “Rational individual behaviour in markets and social choice processes: the discovered preference hypothesis.” In *The rational foundations of economic behavior* (K. Arrow, E. Colombatto, M. Perlaman, and C. Schmidt, eds.), 225–250, Macmillan, London.

Poterba, J.M. (1996), “Retail price reactions to changes in state and local sales taxes.” *National Tax Journal*, 49, 165–176.

Pratt, J.W., D.A. Wise, and R. Zeckhauser (1979), “Price differences in almost competitive markets.” *The Quarterly Journal of Economics*, 93, 189.

Prelec, D., B. Wernerfelt, and F. Zettelmeyer (1997), “The role of inference in context effects: Inferring what you want from what is available.” *Journal of Consumer research*, 24, 118–125.

Rabin, M. and G. Weizsacker (2009), “Narrow bracketing and dominated choices.” *The American economic review*, 99, 1508–1543.

Ratchford, B. (2009), “Consumer search and pricing.” In *Handbook of Pricing Research in Marketing* (V.R. Rao, ed.), Edward Elgar.

Richter, M.K. (1966), “Revealed preference theory.” *Econometrica: Journal of the Econometric Society*, 34, 635–645.

Savage, L. (1954), *The Foundations of Statistics*. Wiley, New York.

Sen, A. (1993), “Internal consistency of choice.” *Econometrica*, 61, 495–495.

Simonson, I. (1989), “Choice based on reasons: The case of attraction and compromise effects.” *Journal of consumer research*, 16, 158–174.

Slovic, P. (1995), "The construction of preference." *American psychologist*, 50, 364.

Sorensen, A.T. (2000), "Equilibrium price dispersion in retail markets for prescription drugs." *Journal of Political Economy*, 108, 833–850.

Spiegler, R. (2011), *Bounded rationality and industrial organization.* Oxford University Press, Oxford.

Stewart, N., N. Chater, and G.D.A. Brown (2006), "Decision by sampling." *Cognitive Psychology*, 53, 1–26.

Stewart, N., N. Chater, H.P. Stott, and S. Reimers (2003), "Prospect relativity: How choice options influence decision under risk." *Journal of Experimental Psychology: General*, 132, 23–46.

Sutton, J. (1992), *Sunk costs and market structure.* MIT press, Cambridge MA.

Tirole, J. (1994), *The theory of industrial organization.* MIT press.

Tversky, A. and D. Kahneman (1974), "Judgment under uncertainty: Heuristics and biases." *Science*, 185, 1124–1131.

Tversky, A. and D. Kahneman (1981), "The framing of decisions and the psychology of choice." *Science*, 211, 453–458.

Tversky, A. and D. Kahneman (1986), "Rational choice and the framing of decisions." *The Journal of Business*, 59, 251–278.

Tversky, A. and I. Simonson (1993), "Context-dependent preferences." *Management Science*, 39, 1179–1189.

Vives, X. (2001), *Oligopoly Pricing: old ideas and new tools.* MIT press, Cambridge, MA.

Wernerfelt, B. (1995), "A rational reconstruction of the compromise effect: Using market data to infer utilities." *Journal of Consumer Research*, 21, 627–633.

Young, D.J. and A. Bielinska-Kwapisz (2002), "Alcohol taxes and beverage prices." *National Tax Journal*, 55, 57–74.

Appendix 1 - Comparison Effect Proofs

Proof of Proposition 1. First I prove it forward. I suppress the history dependence of U for now, and reintroduce it later. Suppose we have two alternatives $a, b \in \mathcal{X}$, with $b \geq a$, and we define $\Delta = b - a$. Then for any $0 \leq \lambda \leq 1$, $WTP_j(a + \lambda\Delta, a)$ can be written as:

$$U(a) = U(a_j + \lambda\Delta_j - WTP_j(a + \lambda\Delta, a), a_{-j} + \lambda\Delta_{-j})$$

Taking the total derivative with respect to λ and WTP,

$$0 = \sum_i \Delta_i \frac{\partial U(\phi(\lambda))}{\partial x_i} d\lambda - \frac{\partial U(\phi(\lambda))}{dx_j} dWTP$$

where $\phi(\lambda) = (a_j + \lambda\Delta_j - WTP_j(a + \lambda\Delta, a), a_{-j} + \lambda\Delta_{-j})$. Rearranging,

$$\begin{aligned} \frac{dWTP(a + \lambda\Delta, a)}{d\lambda} &= \sum_i \Delta_i \frac{\partial U(\phi(\lambda))/\partial x_i}{\partial U(\phi(\lambda))/\partial x_j} \\ &= \sum_i \Delta_i MRS_{i,j}(\phi(\lambda)) \end{aligned}$$

We can now use this path for a particular utility function, denoting $\phi^B(\lambda)$ to denote the path generated from history H^B, and we can evaluate the utility difference between two points under a different history, H^A (I use $U^A(\cdot)$ as a shorthand for $U(\cdot, H^A)$):[32]

$$\begin{aligned} U^A(b_j - WTP_j^B(b, a), b_{-j}) - U^A(a) &= \int_0^1 \left(\sum_{i=1}^n \Delta_i U_i^A - U_j^A \frac{dWTP^B}{d\lambda} \right) d\lambda \\ &= \int_0^1 \left(\sum_{i=1}^n \Delta_i U_i^A - U_j^A \sum_i \Delta_i MRS_{i,j}^B(\phi) \right) d\lambda \\ &= \int_0^1 U_j^A \left(\sum_{i=1}^n \Delta_i (MRS_{i,j}^A - MRS_{i,j}^B) \right) d\lambda \end{aligned}$$

This secures our result: because $\Delta \geq 0$, if $MRS_{i,j}^A < MRS_{i,j}^B$, then $U^A(b_j - WTP_j^B(b, a), b_{-j}) > U(a)$, i.e. evaluated under A the payment WTP^B is too small, so by monotonicity $WTP_j^A > WTP_j^B$.

It also establishes the reverse. Suppose that

$$WTP_j^A(y, x) < WTP_j^B(y, x)$$

[32] In the following I omit the argument $\phi^F(\lambda)$ from the expressions on the right-hand side for U, MRS, and WTP.

and there exists some point $x \in \mathcal{X}$ with

$$MRS^A_{j,i}(x) \geq MRS^B_{j,i}(x)$$

then there must exist a sufficiently small $\Delta_i > 0$, with $y_i = x_i + \Delta_i$ and $y_j = x_j$, $j \neq i$ such that

$$WTP^A_j(x, y) \geq WTP^B_j(x, y)$$

contradicting the assumption, thus proving the proposition. $\square$

Proof of Proposition 2. The optimal demand will satisfy this first-order condition

$$MRS_{j,i}((x_i + q, x_j - p \cdot q, x_{-i,j}), H) = 1/p$$

Because U is quasiconcave, its MRS must be diminishing in every argument. Thus if a change in H causes $MRS_{j,i}$ to increase, q must fall, i.e. demand must fall. $\square$

Proof of Corollary 1. If H^B is a translation of H^A along the price dimension, then $MRS^B_{m,q} < MRS^A_{m,q}$. Each member of the MPL will be of the form $\{(x_q, x_m), (x_q + q, x_m - p_k)\}$, thus from the proposition above, the lower MRS can only cause customers to substitute towards $(x_q + q, x_m - p_k)$. When $H^B$is a translation of H^A along the quantity dimension, the effect on MRS is the opposite, and the result follows analogously. $\square$

Proof of Corollary 2. Because p' first-order stochastically dominates p, we can construct a sequence of price lists $(p^1, ..., p^{m+1}) \in (\mathbb{R}^m)^{m+1}$ with $p^i_j = p'_j$ for $j < i$, and $p^i_j = p_j$ for $j \geq i$. Note that $p^1 = p$ and $p^{m+1} = p'$. The history generated by an MPL based on each element in the sequence is thus a positive translation of the history based on the preceding element, and therefore at each step there is a monotonic effect on the WTP. $\square$

Proof of Corollary 4. We know that if $x \notin c(A, H^A)$ and $y \notin c(A, H^A)$, then $WTP_i(y, x, H^A) \geq 0$, and because H^B is a translation of H^A along the x dimension,

$$WTP_i(y, x, H^B) \geq WTP_i(y, x, H^A)$$

thus x cannot be chosen from H^B, i.e. $x \notin c(A, H^B)$. $\square$

Proof of Proposition 3. Because of our sufficient statistic assumption (Assumption 2) we know that there exists some $H^* \in \mathcal{H}$, with $A^L, A^H \in H^*_T$, and $f(H^L) = f(H^*)$. Then the antecedent of the proposition implies

$$(x^A_i + \delta, x^A_{-i}) \in c(A', H^*)$$

i.e.

$$U((x^A_i + \delta, x^A_{-i}), H^*) \geq U((x^B_i + \delta, x^B_{-i}), H^*)$$

This implies that

$$WTP_i((x_i^B + \delta, x_{-i}^B), (x_i^A + \delta, x_{-i}^A), H^*) \quad \leq \quad 0$$

History H^H can be expressed as two translations of H^L along dimension i, so

$$MRS_{i,j}(x, H^H) \quad < \quad MRS(x, H^L)$$

Thus applying proposition 1,

$$WTP_i((x_i^B + \delta, x_{-i}^B), (x_i^A + \delta, x_{-i}^A), H^H) \quad \leq \quad WTP_i((x_i^B + \delta, x_{-i}^B), (x_i^A + \delta, x_{-i}^A), H^*) \leq 0$$

thus

$$(x_i^A + \delta, x_{-i}^A) \in c(A^H, H^H)$$

$\square$

Proof of Proposition 4. Because of Assumption 2, there must exist some H^*, with $MPL(x, q_L, p)$, $MPL(x, q_H, p) \in H_T^*$, and $\forall x \in \mathcal{X}$, $MRS(x, H^*) = MRS(x, H^L)$, thus $WTP(q_L, H^L) = WTP(q_L, H^*)$, and $WTP(q_H, H^L) = WTP(q_H, H^*)$. And by assumption,

$$\frac{WTP(q_H, H^*) - WTP(q_L, H^*)}{WTP(q_L, H^*)} \frac{q_L}{q_H - q_L} = \varepsilon$$

By Proposition 1, bcause H^H can be expressed as a series of translations of H^L along dimension i, $WTP_j(q_H, H^H) < WTP_i(q_H, H^L)$. Thus, as desired,

$$\frac{WTP(q_H, H^H) - WTP(q_L, H^L)}{WTP(q_L, H^L)} \frac{q_L}{q_H - q_L} < \frac{WTP(q_H, H^*) - WTP(q_L, H^*)}{WTP(q_L, H^*)} \frac{q_L}{q_H - q_L} = \varepsilon$$

$\square$

Proof of Proposition 2. First note that, because H^H can be expresses as a series of increasing translations along the price dimension, then for all $x \in \mathcal{X}$,

$$MRS_{p,q}(x, H^H) < MRS_{p,q}(x, H^L)$$

Second, because of our sufficient statistic assumption (Assumption 2) we know that there must exist an $H^*$with $B(p_L, q), B(p_H, q) \in H_T^*$, and for all $x \in \mathcal{X}$,

$$MRS_{p,q}(x, H^L) = MRS_{p,q}(x, H^*)$$

So these are the same utility functions, and for any p,

$$D_{j,i}(p, H^L) = D_{j,i}(p, H^*)$$

By Proposition 2,

$$D_{j,i}(p_H, H^H) < D_{j,i}(p_H, H^*)$$

Thus

$$\frac{D_{j,i}(p_H, H^H) - D_{j,i}(p_L, H^L)}{D_{j,i}(p_L, H^L)} < \frac{D_{j,i}(p_H, H^*) - D_{j,i}(p_L, H^*)}{D_{j,i}(p_L, H^*)}$$

proving the proposition. □

Appendix 2 - Inference Interpretation Proofs

Proof of Proposition 7. With monotonically unknown separable preferences

$$MRS_{i.j}(x) = \frac{\partial U/\partial x_i}{\partial U/\partial x_j} = \frac{E\left[h(\beta_i)\right] f'(x_i)}{E\left[h(\beta_j)\right] f'(x_j)}$$

Because G_i satisfies the monotone likelihood ratio property, the expected value of $h(\beta_i)$ is decreasing in R_i (Milgrom (1981)), thus, as desired,

$$\frac{\partial MRS_{i,j}(x)}{\partial R_i} < 0$$

□

Appendix 3 - Equilibrium Proofs

Proof of Proposition 8. Corresponding to each first-order condition, there is a total derivative (suppressing the arguments to D):

$$\begin{aligned}(dp_i - dc)\beta[D_{ii}] + \beta^2(p_i - c)\sum_{j=1}^{m} dp_j D_{iij} - dp_1\beta^2(p_i - c)\sum_{j=1}^{m} D_{ii1} \\ +\beta\sum_{j=1}^{m} dp_j D_{ij} - dp_1\beta\sum_{j=1}^{m} D_{ij} &= 0\end{aligned}$$

Where we define $D_{iij} = \frac{\partial}{\partial x_j}D_{ii}$. Dividing by dc and rearranging:

$$(\frac{dp_i}{dc} - 1)\beta[D_{ii}] + \beta^2(p_i - c)\sum_{j=1}^{m}[\frac{dp_j}{dc} - \frac{dp_1}{dc}]D_{iij} + \beta\sum_{j=1}^{m}[dp_j - dp_1]D_{ij} = 0$$

It can be seen that this equation is solved exactly when $\frac{dp_1}{dc} = \frac{dp_2}{dc} = ... = \frac{dp_m}{dc} = 1$. Because we are assuming a unique solution to the first-order conditions, this must be the only solution to the set of total-derivative equations. □

Proof of Proposition 9. The first order condition remains

$$\beta(p_i - c)D_{ii}(p) + D_i(p) = 0$$

where demand is

$$\begin{aligned}D^i(p) &= \int_{-\infty}^{\infty}\int_{\prod_{j\neq i} S_{j,i}} f(v)dv_{-i}dv_i \\ S_{j,i} &= (-\infty, \beta p_j + v_i - \beta p_i)\end{aligned}$$

This can be written in terms of β times the margin of each firm, $\beta(p_i - c)$:

$$S_{j,i} = (-\infty, v_i + \beta(p_j - c) - \beta(p_i - c))$$

Thus the first-order conditions will all be satisfied if $d[\beta(p_i - c)] = 0$, for all i. Thus, letting $\beta = \beta(p_1, .., p_n)$, and $\beta_j = \frac{\partial\beta(p_1,...,p_n)}{\partial p_j}$,

$$\begin{aligned}d[\beta(p_1, ..., p_n)(p_i - c)] &= (p_i - c)\sum_{j=1}^{m} dp_j\beta_j + \beta(dp_i - dc) = 0 \\ \frac{dp_i - dc}{p_i - c} &= -\frac{1}{\beta}\sum_{j=1}^{m}\beta_j dp_j\end{aligned}$$

This quantity cannot be negative.[33] This shows that the proportional change in margins is equal for every firm. This also can be written as:

$$\frac{dp_i}{dc} = 1 - \frac{(p_i - c)}{\beta(\bar{p})} \sum_{j=1}^{m} \beta_j(\bar{p}) dp_j$$

Because $\beta_j < 0$ for all j, and markups are always non-negative, this proves the proposition. $\square$

Proof of Proposition 10. If $p_i > p_j$, then

$$\frac{|dp_i - dp_j|}{dc} = \frac{dp_i}{dc} - \frac{dp_j}{dc} = -\frac{(p_i - p_j)}{\beta} \sum_{k=1}^{m} \beta_j dp_k$$

which is greater than zero, because $\beta' < 0$. The argument is analogous for $p_j > p_i$. $\square$

Proof of Proposition 11. We assume a continuum of firms, so that we can stipulate a zero profit condition which holds with equality: profits must equal the fixed cost C. In a symmetric equilibrium the two conditions will be:

$$\begin{aligned} (p-c)F &= C \\ (p-c)\beta f &= F \end{aligned}$$

When F is uniform, then the conditions become

$$\begin{aligned} (p-c)\frac{1}{n} &= C \\ (p-c)\beta &= \frac{1}{n} \end{aligned}$$

Solving these two conditions we get

$$\begin{aligned} n &= (\beta C)^{-1/2} \\ p - c &= C^{1/2}\beta^{-1/2} \end{aligned}$$

Then we get:

$$\begin{aligned} dp - dc &= -dp\left[\frac{1}{2}\frac{\partial \beta}{\partial p} C^{\frac{1}{2}}\beta^{-\frac{3}{2}}\right] \\ \frac{dp}{dc} - 1 &= -\frac{dp}{dc}\frac{\beta' C^{1/2}\beta^{-3/2}}{2} \\ \frac{dp}{dc} &= \frac{1}{1 + \frac{\beta' C^{1/2}\beta^{-3/2}}{2}} > 1 \end{aligned}$$

[33]Suppose $dp_i/dc < 0$, then $\frac{dp_i/dc - 1}{p-c} < \frac{dp_i/dc}{p} < 0$, so the effect on β is larger than the effect on p, which we ruled out.

and

$$\frac{dn}{dc} = -\frac{1}{2}\left(\frac{C}{\beta}\right)^{-1/2}\frac{d\beta}{dc} > 0$$

Because $\frac{d\beta}{dc} = \frac{\partial\beta}{\partial p}\frac{dp}{dc} < 0$. $\square$

Appendix 4 - Description of Drugstore Data

The data is from a Cambridge, Massachusetts branch of a large national chain of drugstores. The store has floorspace of approximately 500 square meters (5400 square feet). Another drugstore of a similar size, belonging to a competing chain, is located directly across the road.

The store sells a variety of products, principally snacks, groceries, beauty products, stationery, and drugs (both over-the-counter and prescription drugs, however I was not able to observe the labels for the latter). The data was collected by individually photographing price labels for 3,582 different products over a period of 5 days in March 2011, from an estimated 6,000 products in the entire store. From each photo I transcribed the product's ID number, cost code, and price. The ID numbers were matched against a list of ID numbers downloaded from the chain's website on April 4th 2011, which contained much comprehensive information about each product. However 25% of the products photographed were not listed on the website.

The cost (i.e., cost to the retailer) was inferred from the cost code, a sequence of letters on the label. Originally I found the cipher used to decode it posted in an online forum. Subsequently I talked to staff at the store, one of whom confirmed that the code represents cost. The patterns of cost are consistent with plausible changes in product composition. For example, all half gallons of milk sell for $2.29, however the cost is increasing in fat content, the costs for 0%, 1%, 2%, and full fat milk are $1.66, $1.72, $1.76 and $1.78 respectively.

Some products were marked both with a regular price and a temporary sale price. For these I recorded only the regular price. If a promotion was subsidised by the manufacturer without updating the cost (in my observation, cost codes were not updated when switching to and from a promotion), then this would lead to incorrect measurement of margins. Note that in the US the FTC regulates former price comparisons, requiring that the former price be one at which "the article was offered to the public on a regular basis for a reasonably substantial period of time."[34]

[34] See http://www.ftc.gov/bcp/guides/decptprc.htm

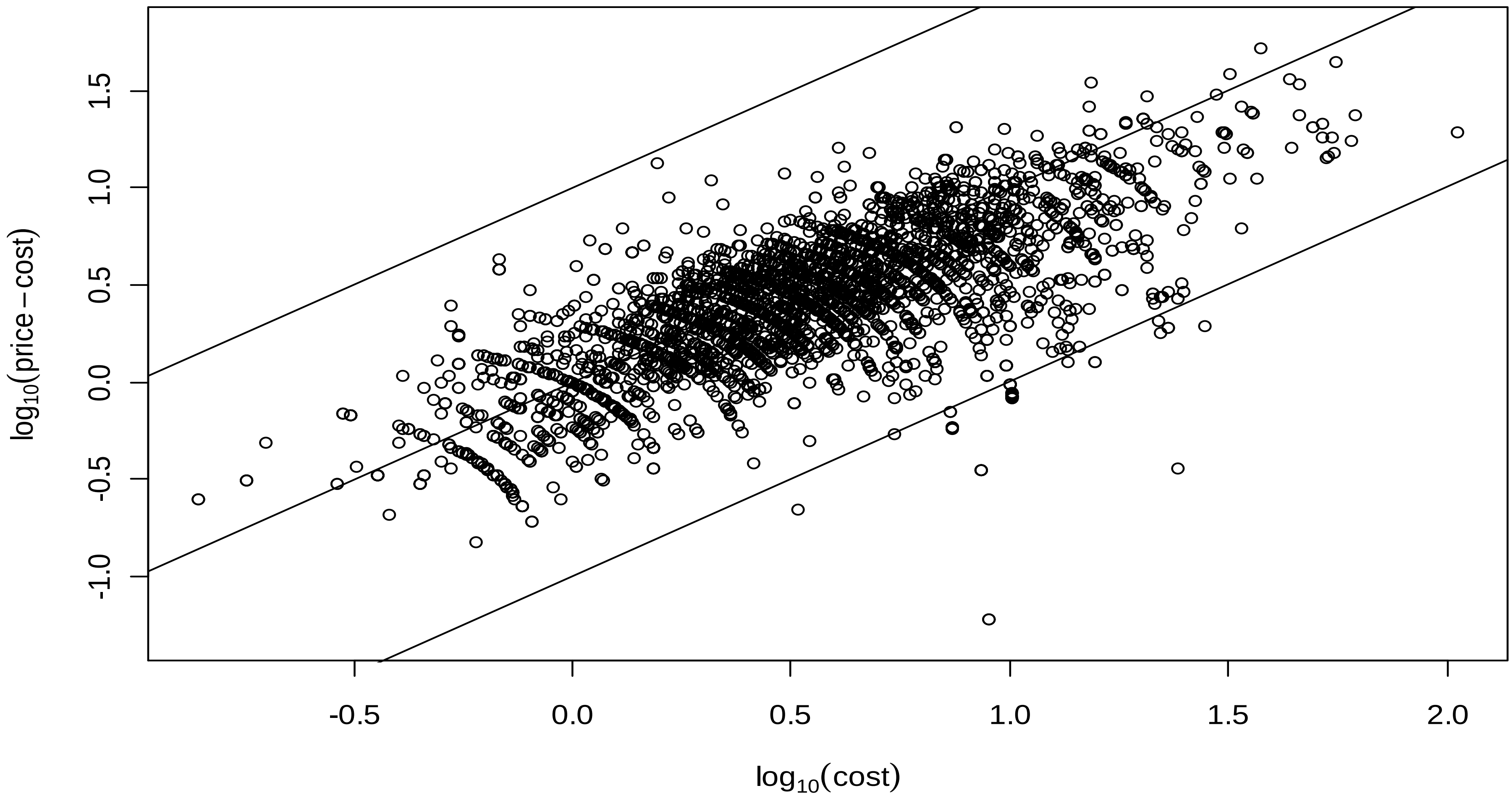


Figure 3: **Cost and Markup in 3,500 Items from a Drugstore.**
Both axes are plotted on a base-10 log scale; the three upward-sloping lines thus represent constant proportional markup rates of 10%, 100%, and 1000%. The curved downward-sloping lines formed by the data points are caused by clustering of prices at common price-points, e.g. 49 cents, 99 cents, etc.

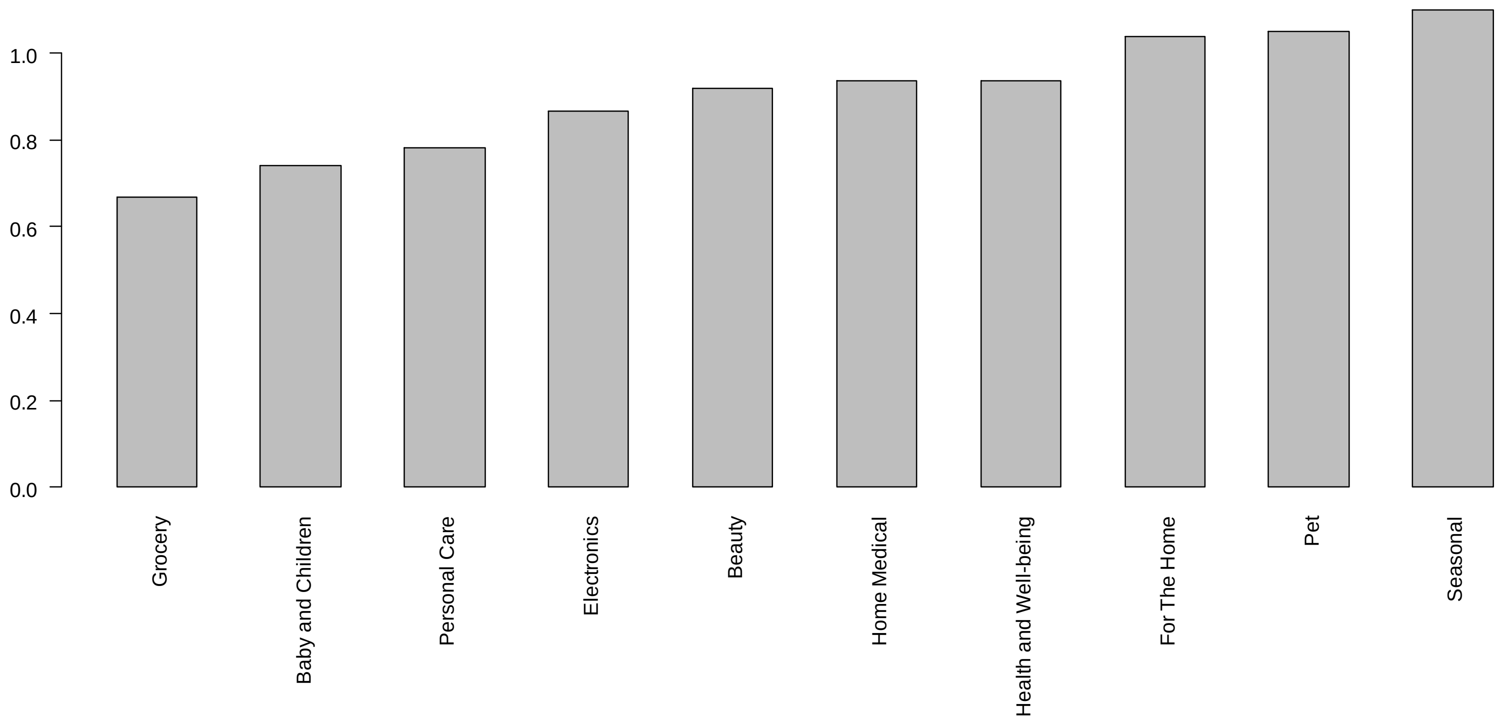


Figure 4: **Average Proportional Markup** ($\frac{p-c}{p}$) **by Product Category**