Essays on Thresholds and on Relative Thinking

Thomas Cunningham

A Thesis Submitted for the degree of Doctor of Philosophy
Ph.D., Economics

Department of Economics
London School of Economics and Political Science

Under the supervision of Professor Francesco Caselli
September 2011

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 24,474 words.

Statement of conjoint work: Chapter 1 of my PhD thesis, titled "Handicapping Politicians: The Optimal Majority Rule in Incumbency Elections", is joint work with Prof. Francesco Caselli of LSE, Inés Moreno de Barreda of Oxford University, and Prof. Massimo Morelli of Columbia University. This work was motivated by Prof. Caselli and Prof. Morelli's idea that an asymmetric electoral rule might be optimal to deal with the well-documented problem of casual incumbency advantage. Inés and myself carried out all the analytical results and writing of this paper with equal share; general decisions about the direction of the paper were made equally between the four authors.

Abstract

The first paper is “Handicapping Politicians: the optimal majority rule in incumbency elections.” The paper examines the incentives of politicians whose position of incumbency allows them to signal their quality through costly effort. Because an election has such a discrete outcome (to win or to lose), politicians will exert the most effort when they expect the election to be close, giving rise to a cluster of signals just above the threshold required to win. Two propositions follow: first, this clustering causes a skew distribution of signals, such that the median expected quality will be above the mean, so more candidates will be re-elected than would be re-elected under full information, i.e. an incumbency advantage. Second, because information is lost when politicians cluster, voters would prefer to handicap the incumbents to make their signals more revealing. In practice this can take the form of a supermajority rule for re-election.

The second paper, “Comparisons and Choice”, shows that many apparently unrelated anomolies of choice can be explained as due to relative thinking. I conjecture that observing larger magnitudes of some goods tends to lower sensitivity to that good. This predicts contrast effects, anchoring effects, scope neglect, and common difference effects. We also introduce a technical novelty, evidence from joint choice, allowing us to map out a utility function.

The third paper, “Relative Thinking and Markups,” applies the model of relative thinking to choice among retailers, when purchasing a single good. We show that demand curves will tend to be less sensitive for higher price goods, so that in the case of unit demand (where markups, price dispersion, and entry are normally independent of cost) markups will be increasing in cost, price dispersion will be increasing in cost, and entry will be increasing in cost. We show that evidence from the IO literature, especially evidence on price dispersion, is consistent with these predictions. We also introduce a novel dataset of costs and prices from a drugstore in which there is a very tight positive relationship between cost and markup.

To my father, Kevin.

Contents

List of Figures

Acknowledgements

With only occasional exceptions the time spent writing this thesis has been a great pleasure for me, and it is very satisfying to finish. There are many people who deserve my thanks.

I give my thanks to Michael Oughtred, for your extremely generous scholarship. I was trying to change ships in midstream, and without your help I doubt I could have made it across. I give my deepest and lifelong thanks.

I give deep thanks to Francesco, for your supervision. Light enough to let me attempt so many things, and heavy enough to bend me towards completion of a few of them.

I give thanks to my other teachers in Economics: Erik Eyster, David Laibson, Andrei Shleifer, Tomasz Strzalecki.

Thank you Ines, for the time we have spent working together, it has been a privilege and a pleasure to have shared so many tricky problems, and to have solved some together.

Thank you to my many friends at the LSE and at the Lorelei and Strand Continental: Maral, Stephan, Amy, Marty, Hande, Christoph, Zsofie, Nathan, Michael, Conrad and Christian, Chris and Stefan, Will and Hester, Carl and Anna, Al and Dave.

Thank you Ma and Sister, for, besides the warmth, and help, and critical and appreciative listening, all the other things.

Thank you Lynn, for being with me.

Part I

Handicapping Politicians

Handicapping Politicians: the optimal majority rule in incumbency elections *

Francesco Caselli, Tom Cunningham, Massimo Morelli, and Inés Moreno de Barreda

June 2011

Abstract

We present a model of electoral competition between an incumbent and a challenger in which everything is symmetric except that voters receive more information about the quality of the incumbent than that of the challenger. The information is received with noise, and is subject to manipulation through costly effort by the incumbent. In equilibrium we show that this model predicts an incumbency advantage, such that incumbents are more likely to be elected than challengers even when their qualities are drawn from identical symmetric distributions, and elected by voters with rational expectations. We also show that a supermajority re-election rule, which sets a threshold for re-election somewhere greater than 50%, improves welfare, mainly through discouraging low-quality politicians from sending high signals.

1 Introduction

In developing countries sitting politicians often are able to exercise considerable influence over the electoral process, and so engineer re-election. This power may explain observations of high re-election rates and voters' desire for term limits. However in developed countries powerful causal incumbency advantages seem to be present in offices which have little direct control over the electoral process, and where there is

*This chapter draws on a joint work with Francesco Caselli, Inés Moreno de Barreda and Massimo Morelli. The idea of the paper was motivated by Francesco Caselli and Massimo Morelli willingness to provide a normative solution to deal with the well-documented problem of casual incumbency advantage. Inés Moreno de Barreda and myself carried out all the analytical results and writing of this paper with equal share; general decisions about the direction of the paper were made equally between the four authors.

no seniority rationale for re-election (Ansolabehere and Snyder; 2002). The principal asymmetry between incumbent and challengers seems to be just in information: that voters are much more informed about incumbents. Cain et al. (1987) state this as follows:

Incumbents win because they are better known and more favorably evaluated by any wide variety of measures. And they are better known and more favorably evaluated because, among other factors, they bombard constituents with missives containing a predominance of favorable material, maintain extensive district office operations to service their constituencies, use modern technology to target groups of constituents with particular policy interests, and vastly outspend their opponents.[1]

This observation is difficult to reconcile with rational expectations, where extra information about the incumbent should not systematically bias voters' beliefs.

We show in this paper that even under rational expectations incumbent power over information can lead to a systematic bias in *election*, such that incumbents are re-elected with a significantly higher probability than in the case without the ability to manipulate the signal. Roughly, the reason is that medium-quality politicians exert a lot of effort to send signals similar to those sent by high-quality politicians, thus generating a skew distribution of signals. The median signal is above the mean signal, meaning that more than 50 percent of signals lead to posterior expected quality that is greater than the average quality, so more than 50 percent of politicians will be re-elected.

More interestingly we show that voters can improve the efficiency of the electoral system by handicapping the incumbent, that is by raising the threshold on expected quality needed to win re-election. A handicap will weaken the incentive of low-quality incumbents to exert effort, while strengthening the incentive of medium and high-quality incumbents to exert effort. The net effect is to raise the average quality of elected politicians.

The handicap we suggest is not time consistent, i.e. voters do not want to enforce it *ex post*. We thus suggest a simple constitutional mechanism for implementation: a supermajority rule, where incumbent politicians require a share of the vote strictly greater than one half in order to win re-election.

In the remainder of this Introduction we discuss related empirical and theo-

[1]See Cain et al. (1987) p10.

retical literature. Section sets up the model, and characterizes the equilibrium. For the rest of the paper we consider a simple symmetric 3-types distribution of politician types. Section analyzes the case of the simple majority rule. Section shows that under the simple majority rule an incumbency advantage exists in equilibrium. Section shows that the optimal re-election rule is a supermajority rule. Section gives an illustrative calibration, and Section discusses related issues and implementation of the supermajority rule.

1.1 Empirical Literature

The observed incumbency advantage is usually thought of as composed of a selection effect and a causal effect. The selection effect is due to incumbents being typically higher quality than challengers, and may be largely benign. A causal effect of incumbency could arise for a number of reasons, good and bad. For example, experience in office may improve the ability of a politician, or make them more effective through earning seniority. On the other hand, the privileges of office may allow them to unfairly influence the next election. The causal and selection effects are difficult to separate, though there are some notable recent attempts.

Levitt and Wolfram (1997) compare repeated pairings of candidates for election to the US Congress, in an attempt to control for the quality of incumbent and challenger. They find that the winner of the previous race has on average a 4% higher vote share in the second pairing.

Ansolabehere et al. (2000) compare county-level vote shares after redistricting in US Congressional elections. They find that incumbents receive 4% fewer votes in counties which have been redistricted into their constituencies, than in counties which remained in their constituency for both elections.

Lee (2008) compares bare winners and bare losers of elections. He finds that a party which barely wins a Congressional election has on average an 8% higher vote share and a 35% higher probability of winning the next election.[2]

Supermajority rules (also called "special majority") are common in constitutions, for example the US Congress can bypass the US President's veto only with a two thirds majority (Goodin and List; 2006).[3] More recently, the Turk-

[2]Note that Lee estimates the incumbency advantage that accrues to the party, not the candidate.
[3]Caplin and Nalebuff (1991) show that under certain conditions, a 64% supermajority voting rule can

ish ruling party narrowly missed the supermajority threshold of two thirds of the seats that would have allowed them to change the Constitution unilaterally. However we are not aware of any supermajority rule being used to handicap the election of incumbents in the way that we suggest.

Finally note that although we know of no explicit incumbent supermajority rules, a similar effect is produced by existing institutions. Constitutional term limits can generally be overturned by an amendment, and constitutional amendments often require a supermajority among legislators. The net effect is then something similar to a supermajority rule on re-election. This is not uncommon in states with term limits, see for example in Colombia in 2004 and in Algeria in 2008.

1.2 Theoretical Literature

The theoretical literature discusses three mechanisms related to incumbency that are relevant to our model. The three mechanisms are (i) that re-election can function as a reward for good behavior, (ii) that extra information about incumbents allows signaling, and (iii) that there can be complementarities between politicians' terms.

An early literature on re-election incentives proposes that voters motivate politician effort by using re-election as a reward for good behavior (e.g. Barro (1973)). In this kind of model a term limit would have an unambiguously negative effect on welfare because it would disable one of the principal mechanisms by which politicians are motivated.

A problem with models of this type is that the threat of punishment is only barely credible. When politicians all have the same quality voters are always indifferent between the incumbent and the challenger. This indifference means that an equilibrium in which voters punish badly behaved politicians can be subgame perfect.[4] However this indifference also means that if there exists any heterogeneity in politician talent, then the equilibrium disappears. For forward-looking voters any difference in perceived talent will dominate incentives to punish or reward sitting politicians. Voters may in fact be partly backward looking (see Smith et al. (1994)), which would of course complicate incentives

eliminate intransitivities in aggregation of preferences.

[4] Along with many other equilibria.

for incumbents. We abstract away from these considerations in this paper.

A second strategic aspect of incumbency exists when politicians differ in ability, and voters observe signals related to ability when a politician is in power.[5] Early papers in this literature include **?**, in which signaling produces a political budget cycle.

Smart and Sturm (2006) use this type of model, a signaling model, to analyze term limits. Their model has two types of politicians and information about the economy which is private to the politician. Re-election incentives cause both types to ignore their private information, for fear of being perceived as a low type. A term limit, which removes re-election incentives, can eliminate the distortion in policy choice and so raise welfare. Under certain parameters their model also predicts that a two-period term limit is superior to a one-period limit, because there will be some sorting of politicians in the first term.

Our model extends this signaling literature, and shows that it has a natural prediction for a rule on vote shares.

Finally, Gerbach (2008, 2009), like us propose a supermajority rule. In Gerbach (2008), the politician's type is fully revealed to voters once they are elected. However low quality incumbents are sometimes re-elected because while in power they implement policies with benefits that are contingent on re-election, thus generating complementarities between terms in office. A supermajority rule can deter such hostage-taking policies. In an independent work, Gerbach (2009) proposes a model in which incumbents signal their ability with costly effort, with similar predictions to ours. Their different model however displays a continuum of pooling and semi-separating equilibria, and hence welfare judgments are derived under assumptions about the likelihood distribution over equilibria. Our modeling assumptions allow us to avoid such equilibrium selection problems.

2 The Model

The game is between two politicians - incumbent and challenger - and a continuum of voters. Both incumbent and challenger are defined by their talent $\theta \in \Theta$. Talent may be understood as the quality of the politician, a character-

[5] This can be seen as a reinterpretation of the intuition in Barro (1973).

istic orthogonal to the political space, valued by every voter in the same way[6]. A few examples of what might be called talent are competency, honesty and charisma. The talents of both politicians are drawn from the same distribution with cumulative distribution function $F(\cdot)$ and mean $\hat{\theta}$.

The asymmetry between incumbent and challenger comes from the fact that during his period in office the incumbent can send a message about his talent to voters. More importantly, the incumbent can boost his message by exerting some costly effort. The message, denoted by $\tilde{\theta}$, is an additive combination of the politician's talent and his effort, $\tilde{\theta} = \theta + e$. The cost of effort is denoted by $c(\cdot)$ and is incurred only by the incumbent.[7]

Voters receive the message with some noise, representing the many unobservables which contribute to political outcomes, and constrain voters' ability to infer a politician's quality. Both the incumbency advantage and the supermajority result can be derived without noise, but noise eliminates pooling equilibria and hence allows us to explore the comparative statics of the equilibria. Also, noise generates a realistically continuous distribution of vote shares, which we use in our calibration.

To differentiate between the information sent by the incumbent and the information received by voters, we will call *message* what the incumbent sends and *signal* what the voters receive. The signal is equal to the original message, plus noise $s = \tilde{\theta} + \epsilon$, where ϵ is drawn from a continuous distribution with mean zero, symmetric and single peaked density distribution function $g(\cdot)$ and cumulative distribution function $G(\cdot)$.

Note that all voters receive the same signal, i.e. the noise is common to all voters.[8] However, voters differ in their preferences for the incumbent. We assume that the utility of voter i given an incumbent with talent θ is given by:

$$u_i(\theta) = \theta + \eta_i \tag{1}$$

where η_i represents voter i's relative preference for the incumbent over the

[6]This concept is also called in the literature *quality* or *valence* (Ansolabehere and Snyder; 2000; Carrillo and Castanheira; 2002; Ashworth and Bueno de Mesquita; 2008a,b).

[7]One might also think of the politician signaling their talent through the choice of a public policy as in Smart and Sturm (2006). We have abstracted of this because we wanted to isolate the informational effect of signalling and its indirect welfare implications.

[8]We discuss in footnote 16 the general effects of heterogenous information.

challenger. We assume that η_i is continuously distributed, with full support on the real line, with cumulative distribution function $H(\cdot)$, density distribution $h(\cdot)$ with both mean and median equal to 0.

This model can be seen as a reduced form of a model in which, after the incumbent sends his message to the population, both incumbent and challenger announce their political platforms. In any subgame perfect equilibrium of such a model, there would be convergence of platforms to the median voter's preferences, and hence the choice of effort is taken as if the voters had preferences given by (1).[9] Finally, we assume that voters support the incumbent when indifferent, though because the noise distribution is atomless, the probability of an indifference occurring is vanishingly small.

Politicians are only office-motivated. Being in office leads to a reward of π. Their only cost is the cost of effort. Thus the incumbent chooses the level of effort to maximize

$$V(\theta, e) = \pi Pr(reelection|\theta, e) - c(e)$$

The game has two decision stages. In the first stage the incumbent sends a message that the voters receive with some noise. In the second stage the voters cast their vote. The outcome of the election depends on the votes cast and the re-election rule. We will denote a re-election rule by q when the incumbent needs at least the fraction q of the votes in order to be re-elected. In Section 3 we consider the particular case of simple majority rule for which $q = \frac{1}{2}$.

Given voters' preferences a simple majority rule is equivalent to giving all power to the median voter, which is in turn equivalent to maximizing a utilitarian social welfare function. On the other hand, as we will discuss later in Section 5, a supermajority rule is equivalent to giving all the power to a voter who is opposed to or dislikes the incumbent. In order to be re-elected the incumbent's talent should be high enough to gain the support from this hostile voter.

[9]There is a recent literature that focuses on the interaction between the choice of effort and the choice of platform (see Ansolabehere and Snyder (2000); Aragones and Palfrey (2002); Carrillo and Castanheira (2002); Ashworth and Bueno de Mesquita (2008a); Meirowitz (2008)) when there is no asymmetry between the candidates. In some of these papers there is divergence of platforms in equilibrium. We abstract from the possibility of divergence and we choose instead to work with a model that corresponds to the more standard convergence outcome because the focus is not on the interaction between valence and political competition.

Given a re-election rule q, an equilibrium is defined by an *effort rule*, $e_q : \Theta \rightarrow [0, +\infty)$ for the incumbent, and a *voting rule*, $v_q : \mathbb{R} \times \mathbb{R} \rightarrow \{0, 1\}$ for the voters such that:

$$\begin{array}{lll} (i) & e_q(\theta) \in argmax_e \quad \pi Pr_\epsilon(reelection|v_q(\cdot), \theta + e, q) - c(e) \\ (ii) & \quad v_q(s, \eta_i) = 1 \quad \textit{if and only if} \quad E[\theta|s, e_q(\cdot)] + \eta_i \geq \hat{\theta} \end{array} \tag{2}$$

where $Pr_\epsilon(reelection|v_q(\cdot), \theta + e, q)$ is the probability of re-election given the voting rule $v_q(\cdot)$, the message $\tilde{\theta} = \theta + e$ and the re-election rule q:

$$Pr_\epsilon(reelection|v_q(\cdot), \theta + e, q) = Pr_\epsilon\left(\int v_q(\theta + e + \epsilon, \eta_i) dH(\eta_i) \geq q\right)$$

and where $E[\theta|s, e_q(\cdot)]$ is the expected talent of the incumbent given that the public signal is s and using a posterior distribution of the incumbent's talent consistent with the equilibrium effort $e_q(\cdot)$.

Finally, we will say that the noise distribution $g(\cdot)$ satisfies the Monotone Likelihood Ratio Property (MLRP) if whenever $\tilde{\theta}_1 > \tilde{\theta}_2$, then $\frac{g(s-\tilde{\theta}_1)}{g(s-\tilde{\theta}_2)}$ increases in s.[10] The MLRP implies that higher signals lead to higher posterior distributions of the talent (higher here meaning first-order stochastic dominance).

The following proposition states than in equilibrium, the incumbent is re-elected whenever the public signal is equal to or above a certain threshold, and is not re-elected otherwise.

Proposition 1 *For any re-election rule q, if the cost of effort, $c(\cdot)$ is strictly convex and the distribution of noise satisfies the MLRP, then in any equilibrium $e_q^*(\cdot)$ and $v_q^*(\cdot)$, the incumbent is re-elected if and only if the public signal is above a threshold k_q, where k_q is given by:*

$$E[\theta|s = k_q, e_q^*(\cdot)] = \hat{\theta} - H^{-1}(1 - q) \tag{3}$$

Equation (3) states that the expected quality of an incumbent who sends a signal $s = k_q$ should equal the expected quality of a challenger, minus the partisan preference of the q^{th} percentile voter towards the incumbent.

The proof of Proposition 1 is similar to that of Theorem 1 in Matthews and

[10]This definition corresponds to the special case of the MLRP defined by Milgrom (1981) when the signal structure is additive.

Mirman (1983) regarding limit pricing. Their setup is close to ours: a monopoly wants to deter the entrant of a possible challenger, and they do so by lowering their price, to signal lower profitability in the market. Analogously, a politician exerts effort to signal their type.

We begin with two preliminary results. In Lemma 1 we show that if the cost function is convex, the message sent by the incumbent is nondecreasing in his type.

Lemma 1 *Given a re-election rule q, if $c(\cdot)$ is strictly convex, and $e_q(\cdot)$ is a best response to $v_q(\cdot)$, then the corresponding message $\tilde{\theta}_q(\cdot)$ is nondecreasing in θ.*

Proof Let $\theta_1 < \theta_2$, and denote $\tilde{\theta}_q(\theta_i)$ by $\tilde{\theta}_i$ and $Pr_\epsilon(reelection|v_q(\cdot), \tilde{\theta}_i, q)$ by $P(\tilde{\theta}_i)$. Since $e_q(\cdot)$ (and therefore $\tilde{\theta}_q(\cdot)$) is a best response to $v_q(\cdot)$,

$$\begin{aligned} \pi P(\tilde{\theta}_1) - c(\tilde{\theta}_1 - \theta_1) &\geq \pi P(\tilde{\theta}_2) - c(\tilde{\theta}_2 - \theta_1) \\ \pi P(\tilde{\theta}_2) - c(\tilde{\theta}_2 - \theta_2) &\geq \pi P(\tilde{\theta}_1) - c(\tilde{\theta}_1 - \theta_2) \end{aligned}$$

Rearranging:

$$c(\tilde{\theta}_2 - \theta_1) - c(\tilde{\theta}_1 - \theta_1) \geq \pi(P(\tilde{\theta}_2) - P(\tilde{\theta}_1)) \geq c(\tilde{\theta}_2 - \theta_2) - c(\tilde{\theta}_1 - \theta_2)$$

Since the distance between the two sets of points is the same: $|(\tilde{\theta}_2 - \theta_1) - (\tilde{\theta}_1 - \theta_1)| = |(\tilde{\theta}_2 - \theta_2) - (\tilde{\theta}_1 - \theta_2)|$, the convexity of $c(\cdot)$ implies that $\tilde{\theta}_1 \leq \tilde{\theta}_2$.$\square$

In Lemma 2 we find sufficient conditions so that each voter's best response is a threshold rule.

Lemma 2 *If $\tilde{\theta}(\cdot)$ is increasing and $g(\cdot)$ satisfies the MLRP, then voter i's best response is a threshold rule:*

$$v(s, \eta_i) = \begin{cases} 0 & if \quad s < k_i \\ 1 & if \quad s \geq k_i \end{cases}$$

where k_i is determined by $E[\theta|s = k_i, e_q(\cdot)] + \eta_i = \hat{\theta}$. Moreover, k_i is decreasing in the preference parameter η_i.

Proof If $\tilde{\theta}(\cdot)$ is increasing in θ and $g(\cdot)$ satisfies the MLRP, the conditional expectation of the talent is increasing in the signal received by the voter (Milgrom; 1981), i.e., if $s_1 < s_2$ then $E[\theta|s_1, e_q(\cdot)] < E[\theta|s_2, e_q(\cdot)]$.

Moreover, since no information is revealed from the challenger, the expected talent of the challenger coincides with the mean of the talent distribution. Therefore, a voter with partisan position η_i supports the incumbent if and only if:

$$E[\theta|s, e_q(\cdot)] + \eta_i \geq \hat{\theta} \tag{4}$$

Since the conditional expectation is increasing and continuous, there is a unique solution k_i to the equation $E[\theta|s = k_i, e_q(\cdot)] + \eta_i = \hat{\theta}$ and voter i follows a threshold rule in which $v(s, \eta_i) = 1$ if and only if $s \geq k_i$. Finally, by the monotonicity of the expectation, k_i is decreasing in η_i. □

Now we prove Proposition 1.

Proof (Proposition 1) For any equilibrium $e^*_q(\cdot)$ and $v^*_q(\cdot)$, if $c(\cdot)$ is convex, Lemma 1 implies that $\tilde{\theta}^*_q(\cdot)$ is nondecreasing in theta. By the MLRP this implies that $E[\theta|s, e_q(\cdot)]$ is nondecreasing in s, and therefore $v^*_q(\cdot, \eta_i)$ is nondecreasing in s. If $v^*_q(\cdot)$ is constant, then $e^*_q(\cdot) \equiv 0$ since the effort is costly and does not change the behavior of voters. But then $\tilde{\theta}^*(\theta) = \theta$ is increasing in θ and Lemma 2 implies that $v^*_q(\cdot)$ is not constant. Therefore $v^*_q(\cdot, \eta_i)$ must be a threshold rule with some threshold $k_{i,q}$. By the monotonicity of the expectation, $k_{i,q}$ is decreasing in η_i. Denote by $\phi_q(\cdot)$ the decreasing function such that $\phi_q(\eta_i) = k_{i,q}$. The set of voters that support the incumbent given a signal s is $S_s = \{i : \eta_i \geq \eta_s\}$ where $\eta_s = \phi_q^{-1}(s)$. Define $\eta_q = H^{-1}(1 - q)$ and $k_q = \phi_q(\eta_q)$. The signal $s = k_q$ is the minimal signal that guarantees reelection under rule q. In effect if $s \geq k_q$ then the share of votes for the incumbent is: $\Pr(\eta_i \geq \eta_s) \geq \Pr(\eta_i \geq \eta_q) = 1 - H(\eta_q) = q$. □

Given a threshold k_q, the probability of re-election for an incumbent that sends message $\tilde{\theta}$ is $\Pr(\tilde{\theta} + \epsilon \geq k_q) = 1 - G(k_q - \tilde{\theta}) = G(\tilde{\theta} - k_q)$, where the last equality comes by the symmetry of the noise distribution.

We can now write the expected payoff of the incumbent as:

$$V(\theta, e, q) = \pi G(\theta + e - k_q) - c(e)$$

then the (local) first and second order conditions for the optimal effort level, $e_q^*(\cdot)$ are:

$$\begin{array}{rcl} \pi g(\theta + e_q^*(\theta) - k_q) & = & c'(e_q^*(\theta)) \\ \pi g'(\theta + e_q^*(\theta) - k_q) - c''(e_q^*(\theta)) & < & 0 \end{array} \tag{5}$$

To guarantee that the local first and second order condition are sufficient for a global optimum we assume throughout the paper the following condition:

$$\inf_e c''(e) > \pi \sup_\epsilon g'(\epsilon) \tag{6}$$

Condition (6) requires the cost function to be sufficiently convex, so that the marginal cost cuts only once the marginal benefit.

Equation (5) together with the definition of the threshold (3), determine the equilibrium.

Up to now we have not made any assumption on the set Θ and the cumulative distribution of the talent $F(\cdot)$. For the rest of the paper we assume that the distribution of the politician talents is symmetric and has three types, i.e. $\Theta = \{\theta_L, \theta_M, \theta_H\}$, with $\theta_H - \theta_M = \theta_M - \theta_L \equiv \delta$ and $p = Pr(\theta_M) = Pr(\theta_L)$.

The symmetry assumption is here because, following the argument of Cain et al. (1987), we want to isolate the effect of the manipulation of the messages on the incumbency advantage. If the distribution of talent was skewed, for example if the median talent was above the mean, we would expect an incumbency advantage even without the manipulation of the messages. To see this, suppose that the voters could perfectly learn a politician's talent once he is in power. Then in more than 50% of elections the voters will discover that the incumbent has greater talent than the expected talent of the challenger, and hence they will strictly prefer to keep that politician. So more than 50% of the candidates will be re-elected. In a symmetric distribution, no incumbency advantage can arise without manipulation of the messages.

We consider three types because it is the simplest model that is rich enough to be able to explain all the mechanism we want to highlight. Simulations with continuous distributions make us believe that the results are true more generally, but we leave derivations for future work.

3 Simple Majority Rule

As a benchmark consider the simple majority rule $q = \frac{1}{2}$. Notice that given the assumptions on the voters' preferences for the incumbent, equation (3) becomes:[11]

$$E[\theta|s = k^*] = \hat{\theta} \tag{7}$$

where k^* denotes the equilibrium threshold in the simple majority case. In other words, the simple majority rule is equivalent to giving all the power to the median voter, the voter that is ex-ante (before receiving the signal) indifferent between the incumbent and the challenger. The incumbent will be re-elected if and only if this voter believes him to have a higher than average talent.

The equilibrium for the simple majority rule has the following properties:

Proposition 2 *With a simple majority rule, the equilibrium is unique. The effort levels satisfy $e_M > e_L = e_H \equiv e^*$ with $e^* = c'^{-1}(\pi g(\theta_H - \theta_M))$ and the threshold signal is given by $k^* = \theta_M + e^*$.*

Proof For clarity we omit the reference to the electoral rule on the equilibrium variables. Given the talent distribution, upon receiving a signal $s = k^*$, equation (7) becomes:

$$\frac{\sum_j \theta_j g(k^* - \tilde{\theta}_j) Pr(\theta_j)}{\sum_j g(k^* - \tilde{\theta}_j) Pr(\theta_j)} = \theta_M$$

and given $Pr(\theta_H) = Pr(\theta_L)$ and $\theta_H - \theta_M = \theta_H - \theta_M$, it simplifies to:

$$g(k^* - \tilde{\theta}_H) = g(k^* - \tilde{\theta}_L) \tag{8}$$

In particular, given the symmetry of the noise distribution, equation (8) implies that the equilibrium threshold will be exactly half-way between the signals sent by the high and low types incumbents:

$$k^* = \frac{\tilde{\theta}_H + \tilde{\theta}_L}{2} = \theta_M + \frac{e_H + e_L}{2} \tag{9}$$

On the other hand, the first order conditions for the equilibrium effort (5)

[11] For clarity we suppress reference to the effort function $e^*_{1/2}(\cdot)$.

together with equation (8) imply that $e_H = e_L$. Denote by e^* this effort level. Then (9) implies $k^* = \theta_M + e^*$.

To see that $e_M > e^*$ notice that, from the single-peakedness and symmetry of g:

$$\pi g(k^* - \theta_M - e^*) > \pi g(k^* - \theta_L - e^*) = c'(e^*)$$

that is, the marginal benefit for an incumbent with type θ_M of exerting effort e^* outweighs the marginal cost of exerting this level of effort. Therefore, $e_M > e^*$.

Finally, replacing $k^* = \theta_M + e^*$ into the first order conditions for e^* given by equation (5), we obtain the equilibrium level e^*:

$$c'(e^*) = \pi g(\delta)$$

where $\delta \equiv \theta_H - \theta_M = \theta_M - \theta_L$ represents the dispersion of the talent distribution[12]. $\square$

The equilibrium can be visualized in Figure 1. Both the talents and the messages can be read on the horizontal axis. The upward sloping lines represent the marginal costs of effort for each type. Equilibrium messages are determined by their intersections with the curve which represents the marginal benefit of exerting effort, $\pi g(\tilde{\theta} - k^*)$. The curve's peak is at $k^* = \theta_M + e^*$, the threshold above which incumbents are re-elected.

The effort level e^* is increasing in π, decreasing in the marginal cost, and decreasing in the dispersion of the incumbent's talent δ. These results are very intuitive, a direct change in the marginal benefit or cost changes the effort level accordingly. Moreover, if the distance between incumbents increases then it is more difficult to fool the voters by exerting effort and therefore the marginal benefits of effort goes down and they exert less effort.

Assuming that the noise is normally distributed with variance σ_ϵ^2 and mean zero, we can further study how the equilibrium effort level changes with the variance of the noise. The change in the equilibrium effort with respect to the variance of the noise depends on the relative size of the variance of the noise

[12]The distance between the talents of the incumbents is a measure of the dispersion of the distribution. In fact the variance of the talents is given by: $Var(\theta) = 2p(\theta_H - \theta_M)^2 = 2p\delta^2$

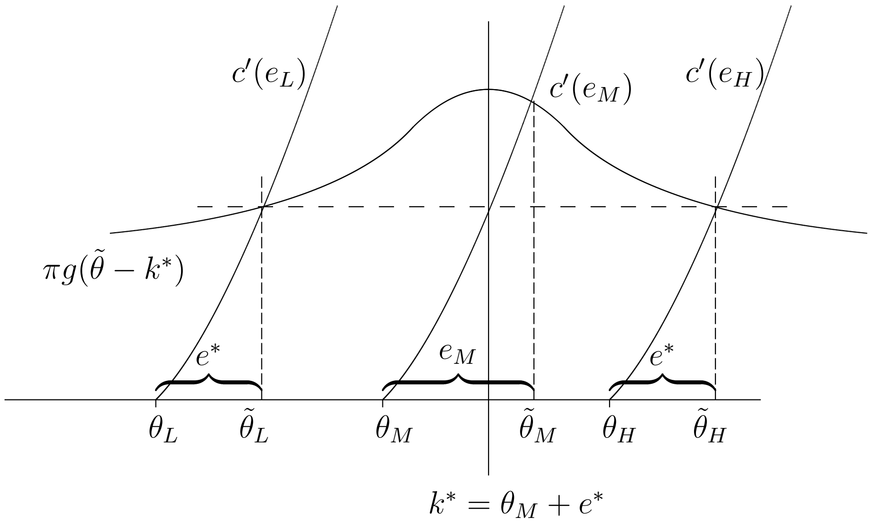


Fig. 1: Equilibrium

and the square of the dispersion of the incumbents:

$$\partial e^* / \partial \sigma_\epsilon^2 < 0 \quad \textit{if and only if} \quad \sigma_\epsilon^2 > \delta^2 \tag{10}$$

To understand this result consider the following two extreme scenarios. Suppose that the signal is *extremely noisy*, then voters do not infer much from the signal and incumbents exert very little effort. If the variance of the signal decreases making the signal more informative, then re-election will be more responsive to the signal received and incumbents will exert more effort. On the other hand, if the signal is *very precise*, incumbents are not going to be able to fool the voters and exert little effort. Condition (10) says that whether we consider the signal *extremely noisy* or *very precise* depends on the relative variances of the two distributions.

4 Incumbency Advantage

One interesting feature of the equilibrium is that the incumbents with middle talent are the ones that exert higher effort. The reason is that the equilibrium threshold is closer to their types and hence they have greater incentive to exert effort.

This extra effort from the incumbents with middle talent implies that the distribution of the messages, signals, and ultimately of expected types, will be negatively skewed (median is above the mean) leading to our result of an incumbency advantage.

We say that there is an incumbency advantage if the expected probability of being re-elected for an incumbent is greater than 50%.

Proposition 3 *The electoral competition model with the simple majority rule exhibits an incumbency advantage.*

Proof From Proposition 2, $e_M > e^*$. The probability of re-election for an incumbent with talent θ_j that sends message $\tilde{\theta}_j$ is then:

$$Pr(reelection \mid \tilde{\theta}_j) = Pr(\tilde{\theta}_j + \epsilon > k^*) = 1 - G(k^* - \theta_j - e_j)$$

The unconditional probability of re-election is therefore:

$$\begin{aligned} Pr(s \geq k^*) &= p(1 - G(k^* - \tilde{\theta}_H)) + p(1 - G(k^* - \tilde{\theta}_L)) + (1 - 2p)(1 - G(k^* - \tilde{\theta}_M)) \\ &= p + (1 - 2p)(1 - G(k^* - \tilde{\theta}_M)) \\ &> \tfrac{1}{2} \end{aligned} \tag{11}$$

Where the second equality follows because $G(k^* - \tilde{\theta}_H) = 1 - G(k^* - \tilde{\theta}_L)$ and the inequality because $e_M > e^*$ so $\tilde{\theta}_M > k^*$. $\square$

Intuitively, when the median voter chooses whether to reappoint the incumbent or not, she compares her updated belief about the talent of the incumbent with the average talent of the challenger. In doing so, she can ignore middle type incumbents because they have just average talent, and hence taking into account the equilibrium messages of the incumbents, the threshold signal would be just the middle point between the messages sent by the low and the high signals. But given that the incumbents with middle talent exert more effort than the others, the message $\tilde{\theta}_M$ will exceed the threshold and therefore they will be re-elected more than half of the times. Because the average re-election probability for low and high types, when combined, is equal to exactly 50%, the total expected re-election probability will be greater than 50%.

5 Supermajority

In this section we consider the social planner's problem of maximizing the total welfare of the voters by choosing a re-election rule (we ignore the utility of the incumbent when computing the social welfare). We prove that the simple majority rule is suboptimal and that the welfare maximizing rule must be a supermajority rule ($q > \frac{1}{2}$).

We proceed in two steps. First, for the simple majority rule equilibrium we show that the voters would be better off if they could commit to a higher threshold to re-elect the incumbent. This commitment is not credible because ex post it is efficient to re-elect the incumbent if the updated beliefs indicate that he is above average (i.e., if the median voter would prefer him). We then propose a way to implement this commitment by setting a supermajority rule that takes decision power from the median voter and gives it to a voter with a partisan position somewhat against the incumbent.

Proposition 4 *In the electoral competition model with the simple majority rule, the welfare maximizing threshold is above the equilibrium threshold k^*.*

Proof Given a threshold k, the expected welfare can be expressed as the value of the outside option (the expected value of a challenger, θ_M), plus the expected change in value from retaining the incumbent:[13]

$$\begin{aligned} EW &= \theta_M + pPr(\tilde{\theta}_H + \epsilon \geq k)(\theta_H - \theta_M) + pPr(\tilde{\theta}_L + \epsilon \geq k)(\theta_L - \theta_M) \\ &= \theta_M + p\delta(G(\tilde{\theta}_H - k) - G(\tilde{\theta}_L - k)) \end{aligned} \tag{12}$$

The optimal threshold is then determined by the first order condition:

$$\frac{\partial EW}{\partial k} = p\delta\left(g(\tilde{\theta}_H - k)(\frac{\partial e_H}{\partial k} - 1) - g(\tilde{\theta}_L - k)(\frac{\partial e_L}{\partial k} - 1)\right) = 0 \tag{13}$$

At the equilibrium threshold, $g(\tilde{\theta}_H - k^*) = g(\tilde{\theta}_L - k^*)$, therefore if we evaluate the derivative (13) at k^*, the direct effect on welfare of a change in the threshold is zero. However, the change in the threshold also affects the choice of effort. Recall that the optimal level of effort given a threshold k satisfies the following

[13] The partisan preferences (η_i) disappear from this expression, because of their zero mean.

first and second order conditions:

$$\begin{aligned} \pi g(\theta_j + e_j - k) &= c'(e_j) \\ \pi g'(\theta_j + e_j - k) - c''(e_j) &< 0 \end{aligned} \tag{14}$$

In particular, totally differentiating the first order condition with respect to k and rearranging:

$$\frac{\partial e_j}{\partial k} = \frac{\pi g'(\theta_j + e_j - k)}{\pi g'(\theta_j + e_j - k) - c''(e_j)} \tag{15}$$

and using the second order condition and the fact that $g'(\theta_L + e_L - k^*) > 0 > g'(\theta_H + e_H - k^*)$ we have that:

$$\left.\frac{\partial e_H}{\partial k}\right|_{k=k^*} > 0 \quad and \quad \left.\frac{\partial e_L}{\partial k}\right|_{k=k^*} < 0$$

Hence, plugging this into (13), the indirect effect on welfare of a raise in the threshold is positive. Increasing the threshold causes θ_H to exert more effort[14] while θ_L will reduce his effort, leading to more separation between the incumbents' signals and as a result an increase in welfare.

We have shown that welfare is improved by marginally increasing the threshold from its Nash equilibrium level. However this is not sufficient to show that a threshold higher than the Nash equilibrium threshold is optimal, because the welfare function may not be single-peaked. We therefore demonstrate below that for any threshold $k < k^*$ the correspondent welfare is strictly lower than the welfare at the equilibrium threshold k^*. To see that, first notice that given θ_j, the optimal effort level e_j defined by equation (14) is a single-peaked function of the threshold k. For a given θ_j, the effort $e_j(\cdot)$ is increasing for $k < \theta_j + c'^{-1}(\pi g(0))$ and decreasing otherwise. Moreover, given equation (14), we have the following identity:

$$e_L(k - (\theta_H - \theta_L)) \equiv e_H(k)$$

so the optimal effort function of the low type is a horizontal shift to the left of the effort of the high type (see Figure 2).

At the equilibrium threshold, $e_L(k^*) = e_H(k^*) \equiv e^*$ so $e_L(k^*) = e_L(k^* -$

[14]A marginally higher threshold also leads the middle θ_M to exert more effort. To see this observe that $\theta_M + e_M - k^* = e_M - e^* > 0$ and hence $g'(\theta_M + e_M - k^*) < 0$ and $\frac{\partial e_M}{\partial k}|_{k=k^*} > 0$.

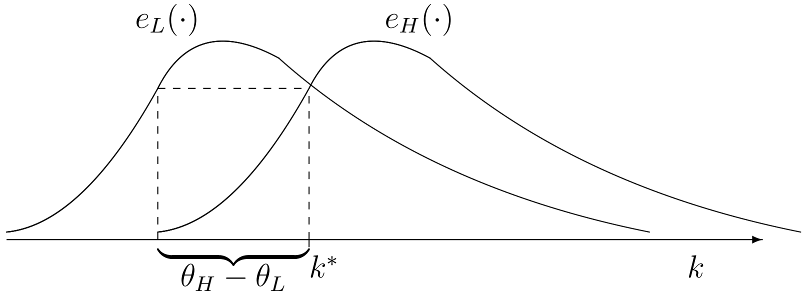


Fig. 2: Effort functions

$(\theta_H - \theta_L))$ which implies that k^* is on the downward-sloping part of curve $e_L(\cdot)$ and on the upward-sloping part of $e_H(\cdot)$. A representation of the effort functions can be seen in Figure 2.

Consider $k < k^*$, then $e_L(k) > e_H(k)$ and hence the distance between the high and low messages under threshold k is smaller than under threshold k^*:

$$\tilde{\theta}_H(k) - \tilde{\theta}_L(k) < \tilde{\theta}_H^* - \tilde{\theta}_L^* \tag{16}$$

Notice that by the symmetry of the noise distribution, the following two remarks are satisfied:

R1: Whenever two points are at a fixed distance h, $G(x) - G(x-h)$ is maximized at $x = \frac{h}{2}$, that is, when the two points are equidistant to the mean.[15]

R2: Given two points equidistant to the mean, the difference in the cumulative distribution is increasing in the distance between the two points:

$$\frac{\partial}{\partial h}\left[G(\frac{h}{2}) - G(-\frac{h}{2})\right] = \frac{1}{2}(g(\frac{h}{2}) + g(-\frac{h}{2})) > 0$$

We can now conclude that for any threshold $k < k^*$ the welfare under

[15]To see this consider the first order condition with respect to x: $g(x) - g(x-h) = 0$, and by the symmetry of $g(\cdot)$, this implies $x = -(x-h)$ or $x = \frac{h}{2}$.

threshold k is lower than under the equilibrium threshold k^*:

$$\begin{aligned} EW(k) &= \theta_M + p\delta(G(\tilde{\theta}_H(k) - k) - G(\tilde{\theta}_L(k) - k)) \\ &\leq \theta_M + p\delta(G(\tfrac{\tilde{\theta}_H(k)-\tilde{\theta}_L(k)}{2}) - G(-\tfrac{\tilde{\theta}_H(k)-\tilde{\theta}_L(k)}{2})) \\ &\leq \theta_M + p\delta(G(\tfrac{\tilde{\theta}^*_H-\tilde{\theta}^*_L}{2}) - G(-\tfrac{\tilde{\theta}^*_H-\tilde{\theta}^*_L}{2})) \\ &= EW(k^*) \end{aligned}$$

where the first inequality follows from R1 and the second from R2 and (16). □

Proposition 4 implies that the voters would be better off if they could commit to re-elect incumbents that have expected talent above a level which is strictly higher than the ex-ante average talent. An increase in the threshold will cause high types to exert more effort and low types to exert less effort. For both types their efforts will not offset the increase in the threshold, so both will be re-elected with a lower probability. But it is the larger fall in the probability of low-type re-election that increases welfare. In other words a supermajority rule makes voters better off entirely through discouraging low quality politicians from seeking re-election.

This higher threshold is not optimal ex post, because it asks the voters not to re-elect some politicians with expected talent strictly greater than the expected talent of the challenger. As discussed in the introduction regarding Barro (1973), it is not clear that individual voters have access to credible commitment devices. However committing to a higher threshold has a natural interpretation with respect to the voters as a whole: a constitutional rule such that incumbents will only be allowed a second term if they exceed some threshold of the vote share strictly greater than one half, i.e. a supermajority rule.

If all voters are identical then this rule, of course, has no effect. However, if the voters differ in their preferences for the incumbent, in the way we have assumed, then a supermajority amendment transfers the decision power from the median voter to a voter that is ideologically opposed to the incumbent.[16]

[16] Another source of voter heterogeneity may be differential information. However if agents are rational, and there is common knowledge of rationality, then it is difficult to argue that the heterogeneous information will not be efficiently aggregated. Information can be indirectly passed through, for example, opinion polls. If a voter compares her own private signal with the aggregated signals of 1000 people in an opinion poll, then the latter would seem to swamp the former. Also voters should vote using the expectations conditional on being decisive; this force will generally make a supermajority rule less effective (see Feddersen

Therefore a supermajority rule acts in effect as a commitment device that sets a higher threshold of talent for re-election.

Proposition 5 *In the electoral competition model, the welfare maximizing re-election rule is a supermajority rule ($q_W > \frac{1}{2}$).*

Proof Given a threshold k, there is a re-election rule that implements that threshold in equilibrium. Denote by $e_k(\cdot)$ the optimal effort the incumbent exerts if he faces threshold k,[17] as a function of hes types. We define $q(k)$ as follows:

$$q(k) = 1 - H(\theta_M - E[\theta|s = k, e_k(\cdot)]) \tag{17}$$

Clearly, setting the re-election rule $q = q(k)$ leads to the equilibrium effort $e^*_{q(k)}(\cdot) \equiv e_k(\cdot)$ and to the equilibrium threshold $k_{q(k)} = k$. To prove Proposition 5 it would be sufficient to prove that $q(k)$ is increasing in k. However this need not be true everywhere. As the threshold gets past a certain point both high and low types will react to an increase in the threshold by lowering their levels of effort (see Figure 2), thus an increase in the threshold could correspond to a lower expected quality from a signal sent at the threshold.

To prove the result we proceed in two steps. First we show that the equation $q(k) = \frac{1}{2}$ has a unique solution at k^*. Then we show that $q(\cdot)$ is strictly increasing at k^*, the equilibrium threshold of the simple majority case. Since $q(\cdot)$ is continuous, this implies that for any $k > k^*$, $q(k) > q(k^*) = \frac{1}{2}$.

Formally, $q(k) = \frac{1}{2}$ if and only if $E[\theta|s = k, e_k(\cdot)] = \theta_M$. By equation (7), k^* satisfies $E[\theta|s = k^*, e_{k^*}(\cdot)] = \theta_M$. To see that k^* is the unique solution to this equation notice that if $E[\theta|s = k, e_k(\cdot)] = \theta_M$, it has to be the case that $\tilde{\theta}_L(k)$ and $\tilde{\theta}_H(k)$ are equidistant to the threshold k. This implies that $e_k(\theta_L) = e_k(\theta_H) = k - \theta_M$. Substituting this in the first order conditions leads to $e_k(\theta_L) = e^* = c'^{-1}(\pi(\theta_H - \theta_M))$ and $k = k^*$.

We now show that $q(k)$ is increasing at k^*, or equivalently, that $E[\theta|s = k, e_k(\cdot)]$ is increasing at k^*:

$$\left.\frac{\partial E[\theta|s = k, e_k(\cdot)]}{\partial k}\right|_{k=k^*} = \left.\frac{\partial}{\partial k}\left[\frac{(\theta_H - \theta_M)p[g(\tilde{\theta}_H - k) - g(\tilde{\theta}_L - k)]}{p(g(\tilde{\theta}_H - k) + g(\tilde{\theta}_L - k)) + (1 - 2p)g(\tilde{\theta}_M - k)}\right]\right|_{k=k^*}$$

and Pesendorfer (1998)).

[17] The effort function $e_k(\cdot)$ solves equation (14).

Denoting by D the denominator of this fraction:

$$\begin{aligned}\left.\frac{\partial E[\theta|s=k,e_k(\cdot)]}{\partial k}\right|_{k=k^*} &= \frac{(\theta_H-\theta_M)p}{D}\left[g'(\tilde{\theta}_H-k^*)(\frac{\partial e_H(k^*)}{\partial k}-1)-g'(\tilde{\theta}_L-k^*)(\frac{\partial e_L(k^*)}{\partial k}-1)\right]\\ &> 0\end{aligned}$$

where the inequality follows because $D>0$, $g'(\tilde{\theta}_H-k^*)=-g'(\tilde{\theta}_L-k^*)<0$ by the equilibrium condition (11) and $\frac{\partial e_j(k^*)}{\partial k}<1$ for $j\in\{H,L\}$ by equation (15).

Therefore, denoting by k_W the welfare maximizing threshold defined by equation (13), $k_W>k^*$ by Proposition 4 and therefore the optimal re-election rule $q(k_W)>\frac{1}{2}$ is a supermajority rule. □

6 Calibration

In this section we do a simple calibration exercise to illustrate the magnitudes involved in our model. We assume that the noise and preference distributions are normal. We also assume a quadratic cost of effort function, with coefficient $\frac{c}{2}$: $c(e)=\frac{c}{2}e^2$, and without loss of generality we set $\theta_M=0$.

The model has five free parameters (discussed below), and we do not estimate all those parameters. Instead we have two modest goals. First, to show that a set of parameters which seem intuitively reasonable (to the authors at least) can reproduce incumbency effects of the right magnitude. Second, to show that the implications for an optimal supermajority rule and its welfare effects are also of an intuitive magnitude.

We target the causal incumbency advantage numbers reported in Lee (2008). That paper uses a regression discontinuity analysis on U.S. Congressional elections, and finds that the difference in the probability of winning an election between a marginal winner and a marginal loser (i.e., a winner or loser of the previous election) is 35%, and that the difference in the average vote share is of 7-8%.[18]

The free parameters of the model are (1) the variance of the noise distribution σ^2_ϵ, (2) the variance of the voters' preferences σ^2_η, (3) the dispersion of the

[18]These numbers correspond to the party rather than the candidate incumbency advantage and average vote share advantage. The problem with the establishment of a candidate incumbency advantage is that there is an endogenous attrition of candidates that distorts the results.

talent distribution $\delta = \theta^H - \theta^M$, (4) the probability of the high and low types p, and (5) the relative cost of effort is c (we normalise $\pi = 1$).[19]

Given the quadratic cost and the normal distributions, the equilibrium of the model is the following:

$$\begin{array}{rcl} e_H = e_L = e^* & = & \frac{1}{c\sigma_\epsilon}\phi(\frac{\delta}{\sigma_\epsilon}) \\ e_M & = & \frac{1}{c\sigma_\epsilon}\phi(\frac{e_M - e^*}{\sigma_\epsilon}) \\ k & = & \frac{1}{c\sigma_\epsilon}\phi(\frac{\delta}{\sigma_\epsilon}) \end{array} \tag{18}$$

where $\phi(\cdot)$ is the standard Normal density distribution.

The probability of winning for an incumbent is given by equation (11) and hence the difference in the probability of winning between the incumbent and the challenger is:

$$x = 2Pr(reelection) - 1 = (1 - 2p)\left(1 - 2\Phi\left(\frac{k^* - e_M}{\sigma_\epsilon}\right)\right) \tag{19}$$

where $\Phi(\cdot)$ is the standard Normal cumulative distribution.

Note that Lee (2008) computes the difference in the probability of winning between a marginal winner and a marginal loser. This avoids the problem of unobserved heterogeneity between winners and losers, if there is sufficient unpredictable noise in votes. Posterior differences between a bare winner and loser thus must be caused by the fact of winning or losing. In our model, all the politicians come from the same distribution of talents and therefore they are ex-ante identical and the difference in the probability of winning comes entirely from having been incumbent.

To compute the average vote share, note that given a signal s the share of voters that support the incumbent is $H(E[\theta|s])$ (see equation (4) for the individual voting rule). Hence the average vote share is given by:

$$AVS = \sum_{j \in \{L,M,H\}} Pr(\theta_j) \int H(E[\theta|s])\, g(s|\tilde{\theta}_j)ds \tag{20}$$

[19] The sufficient condition (6) is translated in the following restriction for the parameters:

$$c \geq \frac{1}{\sigma_\epsilon^2\sqrt{2\pi}}e^{-\frac{1}{2}}$$

and the difference in the average vote share between the incumbent and the challenger is

$$y = AVS - (1 - AVS)$$

We choose the parameters as follows: the difference in quality between θ_H and θ_M is $\delta = 1.4$ (which by symmetry is also the difference between θ_M and θ_L). The probability of the middle type is $(1 - 2p) = 0.7$. The noise has standard deviation $\sigma_\epsilon = 1$, and voters' preferences have standard deviation $\sigma_\eta = \frac{1}{2}$. Finally the cost of effort is $\frac{c}{\pi} = \frac{1}{4}$. These parameters deliver an incumbency advantage matching Lee's estimates, with a difference in the probability of winning of 35% and a difference in the average vote share of 7%.

We can now calculate, using equations (13) and (17), that the optimal supermajority rule is $q_W = 57\%$. This supermajority rule then leads to a welfare increase of 5%, by lowering the proportion of low-quality candidates who are re-elected.

7 Discussion

This paper suggests that if incumbents can use their term in office to influence the voters' perception of their ability, handicapping the incumbent by requiring a higher vote share to be re-elected can improve welfare.

Throughout the paper we have assumed the incumbent faces only a single challenger. To implement our supermajority rule in practice we suggest a two-part ballot: In the first part voters indicate whether they wish to retain the incumbent. In the second part they choose their preferred challenger. This has the advantage of not handicapping the incumbent's party for example, the Republican incumbent can run, and the Republicans can also field a challenger. This ballot structure has been used in some recall elections, e.g. that used for California Governor Gray Davis in 2003.

The type of model we use (screening with noise, in a continuous typespace, but with a discrete reward) is uncommon in the literature. As mentioned earlier, Matthews and Mirman (1983) has the most similar model, though they do not derive analogues of either our incumbency advantage or supermajority results. Besides limit pricing, our approach may be fruitful in a number of other contexts in which thresholds are observed, most naturally entry into jobs which require a minimum score on some test of skill.

A useful extension would be to build a model extending over more than two periods, allowing incumbents to fight multiple elections, and perhaps finding a stationary equilibrium. Unfortunately this is not a simple exercise because two new effects must be modeled: first, the posterior distribution of incumbent types will become asymmetric, through selection; second, voters must now take into account the option value of electing a challenger.

Finally we note that the incumbency advantage and supermajority result can also be shown to occur in a much simpler model with naïve voters. Suppose that there are two kinds of voters - sophisticated and naïve - and that the preferences of the median sophisticated voter coincide with social welfare. Suppose naïve voters always vote for the incumbent because, for example, they are irrationally influenced by advertising, and incumbents always advertise more than challengers. In this case we would expect an incumbency advantage equal to $x\%$ of the vote share, where x is the proportion of voters who are naïve. Further, a supermajority handicap on the incumbent of exactly $x\%$ would make the democratic outcome welfare maximizing. Much of the novelty of this paper is to show that similar results hold even when voters are rational.

References

Ansolabehere, S. and Snyder, J. M. (2000). Valence politics and equilibrium in spatial election models, *Public Choice* **103**: 327–336.

Ansolabehere, S. and Snyder, J. M. (2002). The Incumbency Advantage in U.S. Elections: An Analysis of State and Federal Offices, 1942-2000, *Election Law Journal 1* **3**: 315–338.

Ansolabehere, S., Snyder, J. M. and Stewart, C. (2000). Old Voters, New Voters, and the Personal Vote: Using Redistricting to Measure the Incumbency Advantage, *American Journal of Political Science* **44**: 17–34.

Aragones, E. and Palfrey, T. R. (2002). Mixed Equilibrium in a Downsian Model with a Favoured Candidate, *Journal of Economic Theory* **103**: 131–161.

Ashworth, S. and Bueno de Mesquita, E. (2008a). Elections with platform and valence competition., *Games and Economic Behavior* .

Ashworth, S. and Bueno de Mesquita, E. (2008b). Electoral Selection, Strategic Challenger Entry, and the Incumbency Advantage, *The Journal of Politics* **70**: 1006–1025.

Barro, R. (1973). The control of politicians: an economic model, *Public Choice* **14**(1): 19–42.

Cain, B., Ferejohn, J. and Fiorina, M. (1987). *The Personal Vote: Constituency Service and Electoral Independence.*

Caplin, A. and Nalebuff, B. (1991). Aggregation and Social Choice, *Econometrica* **59 (1)**: 1–24.

Carrillo, J. D. and Castanheira, M. (2002). Platform Divergence, Political Efficiency and the Median Voter Theorem. CEPR Discussion Paper No. 3180.

Feddersen, T. and Pesendorfer, W. (1998). Convinting the Inocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting, *American Political Science Review* **92 (1)**.

Gerbach, H. (2008). Vote-Share Contracts and Democracy. Unpublished.

Gerbach, H. (2009). Higher Vote Thresholds for Incumbents, Effort and Selection. CEPR Discussion Paper No. DP7320.

Goodin, R. E. and List, C. (2006). Special majorities rationalized, *British journal of political science* **36 (2)**: 213–241.

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. House elections, *Journal of Econometrics* **142**: 675–697.

Levitt, S. D. and Wolfram, C. D. (1997). Decomposing the Sources of Incumbency Advantage in the U.S. House, *Legislative Studies Quarterly* **22**: 45–60.

Matthews, S. A. and Mirman, L. J. (1983). Equilibrium Limit Pricing: The Effects of Private Information and Stochastic Demand, *Econometrica* **51 (4)**: 981–996.

Meirowitz, A. (2008). Electoral Contests, Incumbency Advantages, and Campaign Finance, *The journal of Politics* **70**: 681–699.

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

Smart, M. and Sturm, D. M. (2006). Term Limits and Electoral Accountability. CEPR Discussion Paper No. 4272.

Smith, D. J., Dua, P. and Taylor, S. W. (1994). Voters and macroeconomics: Are they forward looking or backward looking, *Public Choice* **78**: 283–293.

Part II

Comparisons and Choice

Comparisons and Choice*

September 30, 2011

Abstract

This paper shows that a wide variety of framing and menu effects can all be derived from a representation of relative thinking, in which increasing a comparator along some dimension will decrease marginal sensitivity along that dimension (i.e., it will lower the marginal rate of substitution). I establish that two well-documented framing effects, the effect of contrast and of anchors, both imply and are implied by this representation of relative thinking. It also implies two menu effects: scope neglect and generalised diminishing sensitivity. I show how menu effects can be identified by using the difference between joint and separate choice. I show that the relative thinking effect would be a rational response to a situation in which relative position is a good proxy for an unobserved true value. However the effect occurs even in experiments which rule out informational effects, thus I suggest the effect can be thought of as a heuristic in the sense of Tversky and Kahneman (1974).

1 Introduction

This paper shows the consequences of comparison effects – also known as relative thinking – on choice behaviour.

*I have benefited greatly from comments by Ruchir Agarwal, Francesco Caselli, Erik Eyster, David Laibson, Andrei Shleifer, Tomasz Strzalecki, Matthew Rabin and seminar participants at Harvard and LSE.

One of the most robust findings from the experimental study of human perception is that judgments of objective quantities (such as length, loudness, brightness) are influenced by comparisons, such that subjects' estimates of magnitude are decreasing in the magnitude of the comparator. Put another way: when an item is in a context in which it is *relatively* small, people tend to judge it to be *absolutely* small.

A natural question is whether this bias affects economic choice through altering judgments of value. This paper shows that a comparison effect, or relative thinking, indeed has predictions for choice in a wide range of settings, and unifies many observed anomalies in choice across various domains.

Analogous to judgments of magnitude, judgments of value can be elicited by asking for willingness to pay, or more generally for a match on another dimension, i.e. asking a question of the form "What value of m would make you indifferent between (x_i, x_j) and (x_i', m)?".[1] Experimental evidence (discussed later) shows that this type of decision is affected by comparator in the same way that perceptual judgments are affected. This paper first establishes that, if preferences are represented with a utility function, then comparators change the marginal rate of substitution between goods. If the magnitude of a comparator along dimension i is represented as c_i, the effect can be summarised as:

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

This expression has a simple graphical representation: because the MRS represents the slope of an indifference curve at some point, an increase in a comparator along the horizontal dimension causes a flattening of a consumer's indifference curves. I show that this representation of comparator effects fits laboratory evidence across a variety of domains and elicitation methods, described variously as sequence effects, anchoring effects, and denominator effects.

I then turn to choice set effects, treating the elements of the choice set themselves as comparators. A difficulty with all theories of choice set effects has been disentangling two different ways in which alternatives can affect decisions: first,

[1] Answers can be made incentive compatible (at least for subjects whose choices obey the axiom of independence) using the system of Becker et al. (1964).

a direct effect through their attributes; and second, an indirect effect through changing the decision-maker's preferences. To solve this problem I introduce evidence from joint choice: i.e., where two independent choice sets are presented side by side, for simultaneous consideration. I argue that violations of revealed preference tend to disappear under a joint choice treatment, meaning that a set of joint choices can be represented as all using the same utility function, thus allowing us to separate the direct and indirect effects of the choice set on decision-making. The model predicts a number of observed choice set effects. Generalising, it predicts that behaviour will appear to exhibit extreme diminishing sensitivity across a variety of elicitation modes, because shifting the choice set out along one dimension will lower sensitivity to that dimension. The principal novel prediction of this model is that diminishing sensitivity will be greatly reduced in joint choice. What evidence there is from joint choice seems to support that prediction.

Turning finally to interpretation, recent research in psychology shows that comparator effects in perception can be elegantly explained as optimal inferences from the environment under plausible constraints. When there is a common unobserved factor which affects all perceptions, then information about relative position can convey information about that unobserved factor, thus rationalising a contrast effect. This, however, does not explain why people still seem to exhibit the same systematic biases when comparators are known to convey no information. I thus suggest that relative thinking can 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.

2 Comparator Effects on Choice

William James described a feature of perception in his *Principles of Psychology* (1890)[2]

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

[2] James (1890)

This effect is known in psychology as a "contrast effect": when 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 experiments rather than introspection, across many domains of perception and judgment.[3] The effect of contrast on judgment is what underlies a whole family of optical illusions: a pair of identical objects look different when each is placed beside a different object; this generates a set of distinct illusions based either on the object's brightness, contrast, colour, or its size.

Turning to economics, similar contrast effects seem to occur in experimental judgments of preference. Day and Pinto Prades (2010) survey sequence effects in the contingent valuation literature: the WTP for a given good tends to fall as the size of the preceding good increases. Sequence effects have also been found in the field, for example when marking exams, judging cases, or meeting potential dates, judgment on the present case tends to be lower if the preceding stimulus was higher (Bhargava (2008)). Bartels (2006) summarises a literature on "denominator effects", in which the willingness to pay (WTP) 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.[4]

Unfortunately none of the experiments in this literature is without flaws. First, the preferences are often not elicited using incentive compatible techniques (e.g. the Becker et al. (1964) mechanism to elicit WTP). In fact most of the experi-

[3]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.

[4]Hsee shows that this is not due to preferences defined over containers: the WTP depends only on the amount of ice cream when judgments are made side by side.

ments asked hypothetical questions. Second, none of these experiments explicitly randomised the magnitude of the comparator. This seems especially likely when subjects judge the value of goods expressed in unfamiliar units. Nevertheless I begin with this somewhat weak evidence for contrast effects to motivate the model in this paper, and I then show that the model can explain a variety of other observed regularities, some of which are better documented than the contrast effect itself.

2.1 Model

The model of relative judgment I introduce in this paper is, under certain assumptions, a unique representation of the contrast effect. The novelty is to allow the utility function to depend on a set of comparators; intuitively, I assume that people make tradeoffs at different, and inconsistent, rates which depend on irrelevant details of the context.

The model predicts the comparative statics with respect to the position of a given comparator, but it makes no prediction about how choice is affected by introducing or removing a comparator.

In this paper I do not specify how to identify the set of comparators in a given situation. The set of comparators can thus be thought of as free variables, or unobservables, when testing the theory. In most of the experimental evidence cited, however, it is relatively clear whether something is a comparator. Dependence on unobservables is a common problem among non-standard theories of decision-making. For example, in reference-point theories, such as prospect theory, predictions differ depending on the reference point.[5] The problem of falsifiability is not so severe for this model: it says that the comparative statics of choice with respect to *any* comparator will be uniform. Thus if any irrelevant comparator affects choice in the wrong direction, the theory is rejected. This differs from reference point theories with unobservable reference points, in which almost any choice can be rationalised by some reference point (Gul and Pesendorfer (2006)).

Let the choice set be $A = \{x^1, .., x^m\}$, with each alternative x^i having at-

[5] Kőszegi and Rabin (2006) give a reference point theory in which the reference point is the lagged rationally expected outcome, thus depending only on the objective attributes of the choice set (plus an assumption about how long the lag is).

tributes along n dimensions, i.e. $x^i = \{x_1^i, .., x_n^i\} \in \mathbb{R}^n$. Then define a super-set of comparators $\mathbb{C} = \{c_1, .., c_n\}$ where each element c_i is a set of comparators along dimension i, labeled $c_i = \{c_i^1, .., c_i^o\}$. I allow choices to depend on the set of comparators, i.e. $U(x|\mathbb{C})$. The common assumption about comparator effects on preferences, which we will use for the rest of the paper, is as follows:

Definition 1. A utility function $U(x|\mathbb{C})$ exhibits *relative thinking* if, for every $1 \leq i, j \leq n$ and comparator c_i^k, with $0 \leq k \leq m$,

$$\frac{\partial}{\partial c_i^k} MRS_{i,j} \equiv \frac{\partial}{\partial c_i^k} \frac{\partial U/\partial x_i}{\partial U/\partial x_j} < 0$$

The claim is that when the comparator shifts out along some dimension then the utility function becomes everywhere less sensitive to that dimension. Figure 1 gives a graphical representation of how choices are affected by the position of a comparator. Because the MRS falls, the indifference curves rotate anti-clockwise to become more parallel with the x dimension. Thus, alternative B's advantage in x becomes less important relative to A's advantage on the y dimension, so in this case choice will change from indifference to strict preference for alternative A.

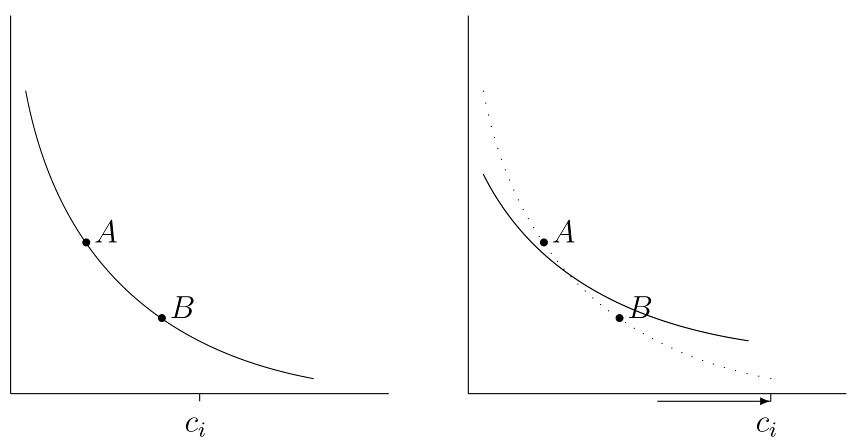


Figure 1: The Effect of a Comparator on Indifference Curves
As the comparator c_i shifts out along the i axis, the MRS$_{i,j}$ everywhere declines, thus rotating indifference curves counter-clockwise.

Some intuitive examples across different domains are as follows. A \$1 difference between alternatives will comes to seem less important when considered in the context of larger amounts of money. A 1% difference in probabilities between alternatives will come to seem less important when considered in the context of larger probabilities. A 1 day delay will come to seem less important when considered in the context of longer delays.

I first show that this representation is not only sufficient but necessary for a contrast effect to occur in matching. Note first how willingness to pay is determined in a utility function. Given an endowment x, then for some increment τ_i along dimension i its **matched value** is equal to an increment v_j along dimension j, i.e.[6]

$$U(x_i, x_j, x_{-i,j}) \quad = \quad U(x_i + \tau_i, x_j - v_j(\tau_i), x_{-i,j})$$

Assuming that U is continuous, strictly increasing, and unbounded in each of its arguments then a unique matched value $v_j(\tau_i)$ will always exist. I will often use $WTP(\tau)$ as a short-hand for $v_j(\tau_i)$, though the results are true for matching on any good, not just for WTP, which is matching on the money dimension.

The theory makes two predictions about the effect of comparators on WTP judgments. First, when valuing some quantity along dimension x, an increase in a comparator along the same dimension will tend to reduce the judged WTP. Evidence for this has already been discussed. However the model also predicts the opposite effect for comparators along the price dimension. When judging WTP a higher price comparator will lower the relative marginal utility of money, thus raising the amount of money required to make the subject indifferent.

Proposition 1. *The following are equivalent:*
(i) the absolute value of the marginal rate of substitution $MRS_{i,j} = \frac{\partial U/\partial x_i}{\partial U/\partial x_j}$ *is everywhere continuous and strictly decreasing in a comparator* C_i *along dimension* i
(ii) the matched value v_j *for a positive increment* τ_i *is continuous and strictly decreasing in a comparator along dimension* i
(iii) the matched value v_j *for a positive increment* τ_i *is continuous and strictly increasing in a comparator along dimension* j

[6] 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.

The new prediction (iii) is supported by a variety of evidence. Best-known is the anchoring effect, where the subject is asked "is this worth more than $\$p$?", and then asked for their willingness to pay. If we consider $\$p$ a comparator along the money dimension, this predicts, as is observed, a higher p is associated with higher reported willingness to pay (the literature is surveyed in Chapman and Johnson (2002), and a good example of a conservative test with strong results is in Fudenberg et al. (2010)).

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 monetary value in the first question is associated with an increased likelihood of accepting the offer in the second question, all else equal.[7]

Finally, it is worth noting that psychological studies also find that comparators along the response dimension tend to have the same effect, of raising the matched value. In the psychology literature this is called an "assimilation effect," as opposed to a "contrast effect" (see Stewart et al. (2003)).

3 The Elements of the Choice Set as Comparators

Suppose we think of the elements of a choice set as themselves comparators, then the model described will immediately make some intuitive predictions.

For example, consider a WTP task, where subjects are asked to state a value for some quantity τ. If we treat the quantity τ as itself a comparator, preferences will be affected by its magnitude, such that subjects asked about a larger quantity will be relatively less sensitive to that good. Nnote that this is not simply diminishing marginal sensitivity; it could be called diminishing *infra*-marginal sensitivity. Thus the model predicts that WTP will grow less quickly than it would without this effect; roughly speaking, the relative-thinking effect will make $WTP(\tau)$ less elastic than it otherwise would be.

[7]The protocol here is choice, rather than matching, but the same comparative statics will hold.

This prediction is obviously difficult to test because it is only a prediction relative to an underlying utility function, which is not directly observable. We cannot know how elastic the WTP function would be without this relative thinking effect. The problem is that changes in the choice set now affect observed choice in two ways: directly, in the usual way, holding preferences fixed; and indirectly, through their effect on the preferences themselves.

The ideal experiment would manipulate the direct and indirect effects independently. This is clearly not possible under the model as it stands. However one further assumption would allow this: if, when making *joint* choices, subjects used the same utility function across each choice set. For example suppose that subjects are asked two questions at once: to state their WTP both for 5 ounces and for 10 ounces of ice cream. Both prompts now serve as comparators, and if the two decisions are made drawing from the same set of preferences then by manipulating one of the choice sets, and observing the effect on choice from the other set, we can separate the direct and indirect effects.

In this section I introduce notation for dealing with joint choice; I give evidence for the assumption necessary to use joint choice as evidence on choice set effects; and finally demonstrate the specific predictions made by this model.

3.1 Context Dependent Choice

The primitive in this model will be choice from a joint choice set, i.e. $c(A|\mathbb{A}) \subseteq A$, where $\mathbb{A}$ is a set of choice sets, and $A \in \mathbb{A}$. I mean that choice is *joint* if the choice sets are offered for simultaneous consideration, but only one choice, randomly selected, is implemented.[8]

The crucial assumption is that the strong axiom of revealed preference (SARP) applies among choices within any joint choice set; i.e., I assume that the choices are acyclic, people do not *simultaneously* revealed prefer a to b, and b to a.[9] This

[8] The joint choice set $\mathbb{A}$ is of course itself a choice set. We could say any choice set A is a *joint* choice between the set of choice sets $\{A^1, ..., A^n\}$ if there exist probabilities $\sum_{i=1}^{n} p^i = 1$ such that $A = (A^1, p^1) \times .. \times (A^n, p^n)$, i.e. the set containing every combination of alternatives from each member choice set, multiplied by the probabilities associated with each set. For example the choice sets $\{A, B\}$ and $\{X, Y, Z\}$ could be combined into the joint choice set: $\{(A, p; X, 1-p), (A, p; Y, 1-p), (A, p; Z, 1-p), (B, p; X, 1-p), (B, p; Y, 1-p), (B, p; Z, 1-p)\}$.

[9] See Richter (1966) and Mas-Colell (1982) for details.

would be violated if, for example, when offered the two choice sets $\{x, y\}$ and $\{x, y, z\}$, you choose x from one, and y from the other. Satisfaction of the SARP guarantees that there exists at least one utility function which rationalizes the joint choices made, thus we can define a utility function $U(x|\mathbb{A})$.[10,11]

If the axiom of independence holds, then of course $c(A^i|\mathbb{A}) = c(A^i)$, because the alternative choice sets do not affect your choice from set i. That axiom rules out menu effects of the kind we are interested in. Some theories of choice weaken the axiom of independence to deal with comparison of outcomes in independent legs of a gamble (e.g. Gul (1991), Bordalo et al. (2010), and Kőszegi and Rabin (2006)) often described as a kind of disappointment-avoidance effect. I ignore that effect here, because it appears that the regret/disappointment effect depends a lot on framing; so that if choices are presented explicitly as separate choice sets, subjects ignore the regret/disappointment effect. This has been the finding of literature on the use of the random-lottery incentive mechanism: Hey and Lee (2005) survey evidence that subjects do not appear to treat a series of choice sets in the same way as one large meta-choice set (this is related to the failure of the axiom of reduction of compound lotteries).[12]

What evidence is there that subjects are consistent when they make choices jointly?

I first discuss framing effects, because choice set dependence can be considered a subset of framing effects.[13] Does the effect of a frame on choice disappear when decisions are made side by side? This is often assumed to be true, but rarely tested. The paper which first documented a variety of many framing effects (Tversky and Kahneman (1986)) says that "the major finding of the present article is that the

[10]Strictly we should write $U(x|\mathbb{C}(\mathbb{A}))$, with $\mathbb{C}(\mathbb{A})$ is the set of comparators in the whole joint choice set. For the remainder of the paper I will leave this implicit.

[11]Spiegler (2011) makes a related point: if framing effects disappear when both frames are presented, then it is difficult to study these effects using the Krepsian paradigm of choice between choice sets.

[12]A related issue is "choice bracketing", described in Tversky and Kahneman (1986) and Rabin and Weizsacker (2009). They ask subjects to choose from both of two choice sets at once. The difference from joint choice as used in this paper, is that both choices were implemented, thus the agents should consider interactions (complementarity, substitutability) between the outcomes.

[13]The formalism can also be easily transposed to pure framing effects, using a choice function like $c(A|F)$, where F represents the frame, as in Salant and Rubinstein (2008); Bernheim and Rangel (2007).

axioms of rational choice are generally satisfied in transparent situations and often violated in non-transparent ones."

I now turn to choice set dependence. The simplest tests would be of this form: when between-subject studies find violations of the weak axiom, e.g. by choosing x from $\{x, y, z\}$ and y from $\{x, y\}$, then an ideal experiment would be to present subjects with the two choice sets jointly, to see if the inconsistency survives. I am not aware of such a direct experiment being run, however we do have evidence from preference reversals. For the case of classic preference reversals (Lichtenstein and Slovic (1971))[14] Ordóñez et al. (1995) found that when subjects are asked to make all three judgments simultaneously (i.e., a choice and two WTA judgments), preference reversals decreased significantly, and were almost eliminated when the game was played for real stakes. 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 significantly as the size of the cup increases, yet when judging payments side by side, both amounts of ice cream are given the same WTP. Similarly Mazar et al. (2009) find that WTP differs that differences in preferences disappear in joint evaluation.[15]

We thus proceed under the assumption that SARP holds in joint choice, allowing us to test theories of menu effects.[16]

4 Members of the Choice Set as Comparators

4.1 Scope Neglect

First I consider how choice set comparators affect behaviour in a matching task; most experimental matching tasks are done with money, so I will denote the

[14] In which a majority of subjects choose one of a pair of bets, but in valuation the average selling price (WTA) is higher for the other bet.

[15] Strictly speaking, in none of these studies did the choices made separately violate the strong axiom of revealed preference, because in a continuous matching task each alternative is unique, so cycles cannot be constructed. However in each case they would violate the strong axiom if we add decisions in which people simply prefer more money to less.

[16] Note that joint choice does not allow us to map out the entire function $u(x|A)$. For example, suppose some choice set A induces only in parts of the attribute space outside of the domain of alternatives in A: such a prediction is untestable.

matching function $WTP(\tau|A)$, although the results are not specific to money.[17]

The model gives a clear prediction for matching: as the magnitude of x increases the utility function will become relatively less sensitive to that dimension, thus diminishing the judgment of willingness to pay. For example, if the underlying WTP function (i.e., the WTP function holding the comparator effect fixed) was a linear function, then the comparator effect would cause it to have decreasing returns to scale, i.e. $WTP(\lambda\tau) < \lambda WTP(\tau)$ for $\lambda > 1$.[18]

As mentioned, elicitation of WTP through a BDM mechanism requires a fairly complex choice set. I abbreviate a Becker-DeGroot-Marschak (BDM) choice set used to elicit the value of τ by denoting it just $A(\tau)$.

Proposition 2. *If the matching function $WTP(\tau|\mathbb{A})$ is linear for a fixed $\mathbb{A}$, and it exhibits relative thinking, then the WTP function will be inelastic,*

$$\frac{dWTP(\tau|\{A(\tau)\})}{d\tau}\frac{\tau}{WTP(\tau|\{A(\tau)\})} < 1$$

Turning to data, there is strong evidence for decreasing returns to scale in matching, though the literature has used different descriptions for this phenomenon in different domains.

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.[19] 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

[17] The relationship between a matching function and a choice set, which I take as primitive, is not trivial. Suppose that matching is done using a BDM mechanism (i.e., a 2nd price auction against a random number), with one hundred one-dollar increments from \$0 to \$99. In this case the choice set contains 100 members each of which is a lottery, for example choosing a \$2 valuation is equivalent to a lottery with four outcomes: a 1% chance each of $(x, \$0)$, $(x, -\$1)$, $(x, -\$2)$, and a 97% chance of $(0, 0)$. This is a complicated choice set, and there is not a unique representation in terms of trading off different goods. For the purposes of this paper I treat the WTP judgment as a joint choice of 100 different choice sets (each implemented with equal probability) of the form (x, p) and $(0, 0)$, then we can represent the outcome along just two dimensions, x and p, and apply the theory directly. I discuss further the problem of alternative representations of a choice set towards the end of the paper.

[18] This can also be described as having local homogeneity everywhere below 1, or having elasticity everywhere below 1.

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

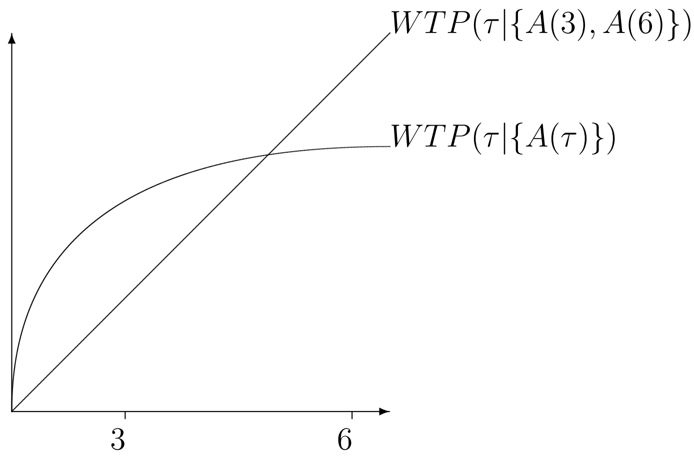


Figure 2: Separate and Joint WTP functions

The graph shows two WTP functions. The concave line represents WTP conditioned just on its argument, τ. The curvature comes from marginal sensitivity to τ everwhere decreasing as τ increases. The straight line represents the utility function evoked by the joint choice set $\mathbb{A} = \{A(3), A(6)\}$. By our assumption any WTP function conditioned on a fixed $\mathbb{A}$ is linear.

saving 2,000, 20,000, or 200,000 birds were \$80, \$78, and \$88, respectively.[20]

Before going on it is important to note that there is an error in the usual description of these effects. They are generally 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.[21] However a better description would be that the evidence shows *low elasticity* in willingness to pay.

There are two ways of explaining the problem. First, if we assume that a zero quantity of some good is always judged to have zero value (zero birds saved, zero probability of harm, etc.) then it is impossible that scope neglect occurs throughout the entire range; if the slope is extremely flat at one point, then it must be correspondingly steep near the origin. Second, the observed slopes of the

[20] 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).

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

WTP functions are not in themselves anomalous. Suppose that the willingness to pay for saving 2,000 birds was \$101 and the willingness to pay for 20,000 birds was \$110 - this would seem to be an anomaly. But if the WTPs were instead \$1 and \$10, although the slope of the function is the same, WTP is now a linear function of quantity, and it does not seem an anomaly.

Thus the slope is not the anomalous part of people's choices, it is the elasticity. Suppose $WTP = \tau^{1/10}$, then when τ increases by a factor of 10, willingness to pay will increase by only $10^{1/10} = 1.25$.

Finally, the theory here predicts that in joint choice, the WTP function should have a much lower elasticity. This is exactly 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 responsiveness to the x variable in within-subjects experiments than in between-subjects experiments.

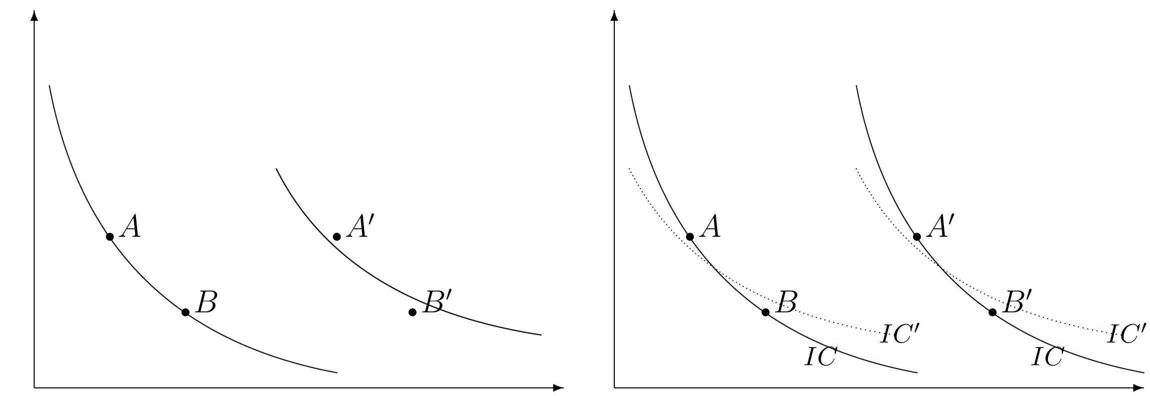
4.2 Common Difference Effects

Tversky and Kahneman (1981) documented strange preferences in purchase decisions: they found that one group of subjects were likely to choose to walk 20 minutes to buy a product for \$10 instead of \$15. However another group of subjects were unwilling to walk 20 minutes to buy a product for \$120 instead of \$125. This is surprising because subjects for whom money enters into their utility functions in a linearly separable way should make the same decision in both cases (and if the utility of money was concave and separable, then they would tend to make the opposite reversal, choosing to walk for the large amount of money and not for the small). A survey of similar evidence can be found in Azar (2007).

The two choice sets can be represented as

$$\{(\$20, 20m), (\$25, 0m)\}$$
$$\{(\$120, 20m), (\$125, 0m)\}$$

The alternatives in the second choice set are identical to those in the first, except for being translated out along the money dimension by \$100 (this is why I have called this phenomenon a "common difference" effect). One interpretation of this



(a) The diminishing sensitivity explanation: both choice use the same utility function, so there must be diminishing MRS along the x dimension.

(b) The relative thinking explanation: each choice set induces a different utility function, but each has parallel indifference curves (i.e., each is quasilinear in money)

Figure 3: Two Models of a Common Difference Effect, where $A \sim B$, but $A' \succ B'$

behaviour is that subjects have diminishing sensitivity to money in this choice; this is the interpretation given in Tversky and Kahneman (1981). An alternative interpretation is given by the model of relative thinking in this paper: treating the two alternatives as comparators, the shift out will lower marginal sensitivity to money, predicting a reversal of preferences as found in the experiment. The two interpretations are illustrated graphically in Figure 2.

Proposition 3. *If preferences are neutral with respect to a common difference along dimension i for any fixed $\mathbb{A}$, i.e. for $\tau > 0$,*

$$(x_1^1, x_2^1) \succeq_{\mathbb{A}} (x_1^2, x_2^2) \iff (x_1^1, x_2^1 + \tau) \succeq_{\mathbb{A}} (x_1^2, x_2^2 + \tau)$$

and the utility function displays relative thinking, then a common difference effect will occur for separate choice, i.e. for $\tau > 0$,

$$\begin{aligned} (x_1^1, x_2^1) \sim_{\mathbb{A}'} (x_1^2, x_2^2) \quad &\Rightarrow \quad (x_1^1, x_2^1 + \tau) \prec_{\mathbb{A}''} (x_1^2, x_2^2 + \tau) \\ \textit{where } \mathbb{A}' \quad &= \quad \{\{(x_1^1, x_2^1), (x_1^2, x_2^2)\}\} \\ \textit{and } \mathbb{A}'' \quad &= \quad \{\{(x_1^1, x_2^1 + \tau), (x_1^2, x_2^2 + \tau)\}\} \\ \textit{and } \ x_2^1 > x_2^2 \end{aligned}$$

Tversky and Kahneman (1981) explain the jacket/calculator results 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 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 $v(5)$[22], which does not change in the two scenarios, and thus would not predict the effect.

The proposition given has a clear prediction: that the diminishing sensitivity will disappear in joint choice. I am not aware of any published experiments which have tried this experiment.

Common difference effects also occur in other domains, usually interpreted as evidence for diminishing sensitivity, though relative thinking may be involved. In inter-temporal 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 amount (Green et al. (1994), and see Frederick et al. (2002) for a survey). In risky choice the common difference effect is also known as the Allais paradox (Allais (1953)): increasing the probabilities of a common prize should not change preferences between gambles, according to expected utility, but it does. 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 remove the 4th bullet, although both actions decrease your chance of death by 1/6.

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

5 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.

5.1 An Inference Interpretation of Relative Thinking

To introduce the inference interpretation, I briefly return to discuss perception. For a long time psychological work on perception and illusions was concerned with giving neurological explanations, describing mechanisms, without explaining why people should be subject to biases which seem easily fixed. Recent work 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. The backgrounds differ either in brightness (lightness contrast illusion), in hue (color contrast), in size (Ebbinghaus illusion), or in contrast (Chubb illusion). In each case the two 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.[23] 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

[23] See Kersten et al. (2004)

share a common multiplicative error,

$$\hat{x} = x/\beta$$
$$\hat{y} = y/\beta$$

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

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

Unobserved common factors occur in many domains of perception, thus generating contrast effects. In an illusion of differing lightness, an unobserved common factor 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 second object. In this way light backgrounds tend to make objects look darker, even for perfectly rational subjects.

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

To summarise, existence of a common unobserved factor can explain contrast effects: because 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's) as known, but the importance of each attribute (β) is a matter of judgment. The following are sufficient conditions for relative thinking:

Definition 2. 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) all xs are known, but βs are unknown to the decision-maker (the DM has a prior over each β)
(iii) the DM believes each comparator $c_i \in \mathbb{C}$ (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[24].

Proposition 4. *If the DM can be described as having monotonically unknown separable preferences then*

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

Note that the decision-maker must take the expectation of utility, because the β's are uncertain. Thus proposition 4 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 we have discussed as anomolies: contrast, anchoring, scope neglect, and common difference.

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

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

This utility function has two interesting properties. First, suppose that utility is linear in money, so 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 (we have already assumed that the utility function

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

is linear in each of its arguments). 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, we 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.

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

5.2 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.[25] 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. Finally, contrast effects occur even when the order of alternatives is random, as in Bhargava (2008).

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, 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 by Tversky and Kahneman (1974), in which simple rules are used which perform well in most typical contexts. Biases occur when they are applied outside that set of contexts.

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

6 Discussion

6.1 Related Literature

The literature on psychology of judgment has a number of strands which deal with context effects on judgment and choice.

First, there is a large experimental literature documenting context influences on judgment across different domains, for an overview see Poulton (1989).

Second there are a class of general theories trying to fit these patterns, the best known is Parducci's range-frequency theory (Parducci (1965),Parducci (1995)) which proposes that the set of stimuli can influence objective judgments of magnitude and of preference. Parducci proposes that judgment of either magnitude or value can be positively influenced by two factors: (1) its position in the range presented (distance between the endpoints), and (2) its percentile rank in the distribution of stimuli. Note that this gives the same predictions for comparator effects as the model presented here, except for changes in the position of the smallest comparator in the set.

Third, treatment of perception as an optimal inference problem is a quite recent research area.[26] For most of the 20th century research in perception tried to relate behaviour to the structure of the nervous system. However since the mid 1990s a large empirical program has emerged modelling perceptual judgment as optimal inference, see for example Kersten et al. (2004) and Purves et al. (2011).

An important related paper is Hsee and Zhang (2010), which proposes an informal theory of "evaluability". They say that subjects are more sensitive to changes in a certain dimension if (i) the evaluation mode is joint rather than separate; (ii) they know more about the domain; and (iii) the dimension is intrinsically evaluable. The model of relative thinking in this paper could be thought of as a foundation for evaluability effects.

Also related is recent work on menu effects, where the weights put on different dimensions depend on the choice set. An early attempt was Tversky and Simonson (1993), more recent are Kőszegi and Rabin (2006), Bordalo (2011), and Koszegi and Szeidl (2011).

[26] Though Helmholtz first stated the problem of perception as an inference problem in the 1860s.

6.2 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 thought of 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 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).

7 Conclusion

This paper has tried to show the connection between two broad phenomena observed in choice: First, choice behaviour in experiments often seems to show rapidly diminishing sensitivity in many domains - in probability, time, money, and in the quantity of a product. This phenomenon holds for both choice and matching. Second, the value put on a good (both in choice and in matching) is influenced by irrelevant comparisons, so that when a quantity comes to be relatively small it is treated as absolutely small. I have tried to show that all these phenomena fit a pattern of preferences being affected by the context, such that higher comparators lower sensitivity.

I have also shown that anomalies often disappear in joint choice, which allows us to define a utility function common to multiple choice sets, and thus observe how choice sets affect preferences.

Finally I suggested an interpretation of the behaviour, as optimal for a common set of situations, in which relative position is a good indicator for absolute value.

References

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. (2007), "Relative thinking theory." *Journal of Socio-Economics*, 36, 1–14.

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.

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

Bernheim, B.D. and A. Rangel (2007), "Toward choice-theoretic foundations for behavioral welfare economics." *American Economic Review*, 97, 464–470.

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.

Bordalo, P. (2011), "Salience theory and choice set effects." Working Paper.

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

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.).

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.

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.

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.

Gul, F. (1991), "A theory of disappointment aversion." *Econometrica*, 59, 667–686.

Gul, F. and W. Pesendorfer (2006), "The revealed preference implications of reference dependent preferences." *manuscript, Princeton University.*

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

Hey, J.D. and J. Lee (2005), "Do subjects separate (or are they sophisticated)?" *Experimental Economics*, 8, 233–265.

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

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., 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.

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.

Koszegi, 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.

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. (1982), "Revealed preference after samuelson." In *Samuelson and Neoclassical Economics* (G.R. Feiwel, ed.), 72–82, Amsterdam: Kluwer.

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

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

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.

Parducci, A. (1965), "Category judgment: A range-frequency model." *Psychological Review*, 72, 407–418.

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

Poulton, E.C. (1989), *Bias in quantifying judgements*. Lawrence Erlbaum Associates, Inc.

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.

Purves, D., W.T. Wojtach, and R.B. Lotto (2011), "Understanding vision in wholly empirical terms." *Proceedings of the National Academy of Sciences*, 108, 15588–95.

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

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

Salant, Y. and A. Rubinstein (2008), "(a, f): Choice with frames." *The Review of Economic Studies*, 75, 1287–1296.

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

Spiegler, R. (2011), "Comments on "behavioral" decision theory." Working Paper.

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.

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.

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

8 Appendix - Proofs

Proof of proposition 1. In what follows I will use an abbreviated utility function, dropping all the elements of the endowment $\{x_1, ..., x_n\}$ which do not change, so that

$$U(\tau_i, -v_j) \equiv U(x_i + \tau_i, x_j - v_j, x_{-i,j})$$

In this proof we wish to compare three utility functions: U, an initial utility function, defined by some set of comparators; U^i, one which differs only by having one comparator along dimension i be greater; and U^j, which differs from U only in having one comparator along dimension j be greater. The matched values, v_j, v_j^i, and v_j^j are defined by the following equations

$$\begin{aligned} U(0,0) &= U(\tau_i, -v_j) \\ U^i(0,0) &= U^i(\tau_i, -v_j^i) \\ U^j(0,0) &= U^j(\tau_i, -v_j^j) \end{aligned}$$

Using this notation the three conditions can thus be represented as

$$\begin{array}{cc} (i) & MRS_{i,j}^i < MRS_{i,j} < MRS_{i,j}^j \\ (ii) & v_j^i < v_j \\ (iii) & v_j^j > v_j \end{array}$$

The last two conditions are in turn are equivalent to the following:

$$\begin{aligned} U^i(0,0) &> U^i(\tau_i, -v_j) \\ U^j(0,0) &< U^j(\tau_i, -v_j) \end{aligned}$$

which say that the matched value v_j, which was indifferent under the original set of comparators, will become too high a price when a comparator increases on the i dimension (under U^i), and will become too low a price when a comparator increases on the j dimension (under U^j). This equivalence holds because the utility functions are continuous and strictly increasing in dimension j.

I will first show that (i) implies (ii) and (iii). The difference in utility between two points on a utility function can be expressed as the integral over the marginal utility along any path between those points. One path is along the indifference curve connecting the endowment point ($\{0,0\}$) and the matched allocation ($\{\tau_i, -v_j\}$),

$$U(\tau_i, -v_j) - U(0,0) \;=\; \int_0^{\tau_i} \left(U_i(t, -v(t)) - U_j(t, -v(t)) \left.\frac{dv}{dt}\right|_{dU=0} \right) dt$$

Here U_i and U_j are partial derivatives, and $v(t)$ is the matched value of t, defined as above. We can thus express this in terms of the marginal rate of substitution between i and j

$$MRS_{i,j} = \frac{U_i}{U_j} = \left.\frac{dv}{dt}\right|_{dU=0}$$

So

$$\begin{aligned} U(\tau_i, -v_j) - U(0,0) \;&=\; \int_0^{\tau_i} \left(U_i(t, -v(t)) - U_j(t, -v(t)) MRS_{i,j} \right) dt \\ &=\; \int_0^{\tau_i} U_j \left(MRS_{i,j} - MRS_{i,j} \right) dt \end{aligned}$$

This last expression is zero, simply showing that the utility at the beginning and the end of an indifference curve must be the same, because utility does not change anywhere along that curve (i.e., the gradient theorem).

However, now consider evaluating the change in utility along exactly the same path according to one of the other utility functions, U^i or U^j. We get these expressions

$$\begin{aligned} U^i(\tau_i, -v_j) - U^i(0,0) \;&=\; \int_0^{\tau_i} U^i_j \left(MRS^i_{i,j} - MRS_{i,j} \right) dt \\ U^j(\tau_i, -v_j) - U^j(0,0) \;&=\; \int_0^{\tau_i} U^j_j \left(MRS^j_{i,j} - MRS_{i,j} \right) dt \end{aligned}$$

The term in brackets is no longer zero. In fact, under condition (i), we know that $MRS^i_{i,j} < MRS_{i,j} < MRS^j_{i,j}$. Thus we can conclude, as desired:

$$\begin{aligned} U^i(\tau_i, -v_j) \;&<\; U^i(0,0) \\ U^j(\tau_i, -v_j) \;&>\; U^j(0,0) \end{aligned}$$

Next we wish to show that (ii) and (iii) imply (i). Suppose a counterexample existed, i.e. suppose that $v_j^i < v_j$ always holds but there is some point $\{x_1, ...x_n\}$ with

$$MRS_{i,j}^i \geq MRS_{i,j}$$

Now consider an expression for a matched value, $v_j(\tau_i)$:

$$U(x_i, x_j, x_{-i,j}) \;\; = \;\; U(x_i + \tau_i, x_j - v_j(\tau_i), x_{-i,j})$$

We can solve for the derivative (omitting the arguments for U and MRS)

$$\begin{aligned} 0 &= d\tau_i \frac{\partial U}{\partial x_i} - dv_j \frac{\partial U}{dx_j} \\ \frac{dv_j}{d\tau_i} &= -\frac{\partial U/\partial x_i}{\partial U/\partial x_j} = -MRS_{i,j} \end{aligned}$$

Thus the matched value $v_j(\tau_i)$ can be implicitly defined as,

$$v_j(\tau_i) = \int_0^{\tau_i} MRS_{i,j}(x_i + s, x_j + v(\tau_i - s), x_{-i,j})ds$$

Thus for a sufficiently small τ_i, $v_j^j > v_j$ if and only if $MRS_{i,j}^j > MRS_{i,j}$. □

Proof of Proposition 2. This follows fairly directly from the preceding proposition. (Note: for compactness I represent the joint choice sets as superscripts, instead of conditioning arguments). We know that $MRS_{x,p}^{\{A(\lambda x)\}} < MRS_{x,p}^{\{A(x)\}}$, and by applying the proposition $WTP^{\{A(\lambda x)\}}(\lambda x) < WTP^{\{A(x)\}}(\lambda x)$. But because each utility function is linear, $WTP^{\{A(x)\}}(\lambda x) = \lambda WTP^{\{A(x)\}}(x)$, so

$$WTP(\lambda x|\{A(\lambda x)\}) \;\; < \;\; \lambda WTP(x|\{A(x)\})$$

as desired. □

Proof of Proposition 3. To show that relative thinking entails a common difference effect, we use the three choice sets as defined in the proposition, which gives us:

$$U(x_1^1, x_2^1|\mathbb{A}') \;\; = \;\; U(x_1^2, x_2^2|\mathbb{A}')$$

$$
\begin{aligned}
U(x_1^1, x_2^1 + \tau|\mathbb{A}') &= U(x_1^2, x_2^2 + \tau|\mathbb{A}') \\
U(x_1^1, x_2^1 + \tau|\mathbb{A}") &< U(x_1^2, x_2^2 + \tau|\mathbb{A}")
\end{aligned}
$$

Where the last step uses proposition 1: the change in MRS when shifting from $\mathbb{A}'$ to $\mathbb{A}$" makes the relative advantage of alternative 1 along dimension 2 become less significant. □

Proof of Proposition 4. With monotonically unknown 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,

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

□

Part III

Relative Thinking and Markups

Relative Thinking and Markups*

September 30, 2011

Abstract

In experiments subjects regularly trade off time and money at inconsistent rates, apparently becoming less sensitive to money when considering the purchase of a more-expensive item. In this paper I introduce a demand function that matches this behaviour, and derive the predictions for equilibrium market structure, and then finally discuss market evidence. The demand function is derived from a utility function which has context-dependent sensitivity to different goods (e.g., time, money). In market equilibrium with unit total demand the model predicts three phenomena: higher cost goods tend to have higher markups; higher cost goods tend to have greater price dispersion; and higher cost goods will have a larger number of sellers. The first two facts are commonly observed in the empirical IO literature. Finally I introduce a novel dataset of 3,500 markups from a branch of a chain drugstore, with cross-section evidence in support of the theory.

1 Introduction

In the study of industrial organisation, assumptions about utility can put important restrictions on demand functions. Assuming a utility function means assuming that consumers trade off different goods at consistent rates.

*I would like to thank Ruchir Agarwal, Francesco Caselli, Erik Eyster, David Laibson, Ariel Pakes, Matthew Rabin, Andrei Shleifer, Tomasz Strzalecki, Glen Weyl and seminar participants at Harvard for extremely helpful comments.

However a famous laboratory experiment seems to show that rates of trade-off between time and money can vary drastically between situations. Tversky and Kahneman (1981) found that most of their experimental subjects were willing to drive 20 minutes to save $5 from a $15 item, but not to drive 20 minutes to save $5 from a $125 item. If the subjects had utility linear in money, then their willingness to drive should be the same in both situations. If instead they had concave utility of money, the preference reversal should go in the opposite direction.

If people are subject to this bias in the real world, i.e. valuing money less when considering larger values, this will have important effects on demand functions, and therefore important effects on the equilibrium distribution of prices and quantities. In this paper I show how this relative thinking will affect demand under fairly general assumptions, deriving three predictions about equilibrium. First, margins will be increasing in cost (equivalently, pass-through will be greater than 1, or demand will be cost-amplifying[1]). Second, price dispersion will be increasing in cost. This holds whether dispersion is measured as the range or standard deviation of prices. Third, entry will be increasing in cost. Each follows simply because higher costs, insofar as they lead to higher prices, lower customers' sensitivity to marginal units of money, having an effect on prices equivalent to an increase in transport costs (more generally, an increase in any costs of substitution between goods).

Much empirical literature supports these predictions. For the markup predictions, most time series studies find pass-through rates greater than 1. For the dispersion predictions we only have cross section studies, though again these strongly support the predictions. I discuss in detail other explanations for these patterns.

In this paper I also introduce a novel dataset of 3,500 costs and prices from a branch of a chain drugstore. The dataset is unusual in having cost and price, and therefore markups. The dataset shows an extremely close relationship between cost and markup (here defined as price minus cost). One striking feature is how few outliers there are. For example, there are around 400 products with a cost (to the retailer) of more than ten dollars, and around 400 with a cost of less than one dollar. Seventy percent of the former have a markup of more than $5. Yet, of the

[1] Weyl and Fabinger (2009)

latter, not a single product has a markup of more than \$5. Aside from relative thinking, differences in markups can be caused by many factors: elasticity of demand, customer demographics, frequency of purchase. But it is remarkable that there should not be a single \$1 item with the right mix of elasticity, demographics, and purchase frequency to justify a \$5 markup.

2 Relative Thinking

Tversky and Kahneman (1981) introduced relative thinking into the literature on biases in economic decision-making with this pair of questions, given to two different groups of subjects:

> (A) Imagine that you are about to purchase a jacket for \$125, and a calculator for \$15. The calculator salesman informs you that the calculator you wish to buy is on sale for \$10 at the other branch of the store, located 20 minutes drive away. Would you make the trip to the other store?
>
> (B) Imagine that you are about to purchase a jacket for \$15, and a calculator for \$125. The calculator salesman informs you that the calculator you wish to buy is on sale for \$120 at the other branch of the store, located 20 minutes drive away. Would you make the trip to the other store?

Someone who makes consistent trade-offs between time and money should answer the same way to both questions. However Tversky and Kahnemann found that in case A, when the \$5 discount is off a \$15 total, then 68% of subjects said they will make the trip to the other store. In case B, when the \$5 discount is off a \$125 total, only 29% of subjects were willing to make the trip.

The effect has been replicated in many variations, though mostly in hypothetical experiments. As an illustration of the magnitude of these hypothetical responses, subjects in Azar (2011) 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.[2]

There is not an agreement in the literature on how to describe the source of this inconsistency. Tversky and Kahneman (1981) say it comes from diminishing sensitivity and narrow bracketing. Azar (2008) says that higher prices tend to lower psychic transport costs. In a companion paper (Cunningham (2011)), I give an alternative explanation: tradeoff decisions are influenced by the choice set, such that observing higher magnitudes along any dimension will tend to lower sensitivity to that dimension. Thus higher prices cause lower sensitivity to money, leading to demand which is less sensitive for high prices than for low prices, thus higher markups for high cost goods.

3 Modeling Relative Thinking

I here introduce a model of decision-making where the utility function depends on the choice set in a particular way. The model is discussed in greater detail in Cunningham (2011).

Intuitively, the model states that subjects become less sensitive to a marginal difference when they are confronted with larger magnitudes of that good. Thus when they are considering large amounts of money, a small difference comes to seem less important. So for more expensive goods, consumers are less likely to take a trip to save some money.[3]

Let subjects choose from a choice set A containing m alternatives $\{x^1, .., x^m\} = A$, where each alternative is a vector of n attributes $x^i = \{x^i_1, .., x^i_n\} \in \mathbb{R}^n$. An ordinary utility function is just a function of the attributes of each alternative $U(x_1, .., x_n)$, whereas we here instead consider a choice-set dependent utility func-

[2]The effect is so strong, it seems to sometimes affect economists writing on markups. For example Lach (2002) explains some data as consistent with rationality because “search costs are low relative to the high price of the good and, as a consequence, more searching for the lowest price is undertaken.”

[3]Incidentally 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.”

tion $U(x_1, .., x_n|A)$.

Definition 1. A choice-set dependent utility function $U(x|A)$ exhibits *relative thinking* if, for every $1 \leq i, j \leq n$ and alternative $0 \leq k \leq m$,

$$\frac{\partial}{\partial x_i^k} MRS_{i,j} \equiv \frac{\partial}{\partial x_i^k} \frac{\partial U / \partial x_i}{\partial U / \partial x_j} \leq 0$$

The formula expresses that an increase in the magnitude of any of the alternatives' attributes along dimension i will cause the relative sensitivity to everywhere decrease, i.e. the marginal rate of substitution will get smaller. Figure 1 shows that, when B shifts out along the horizontal dimension, this causes marginal utility with respect to that dimension to decrease, and therefore causes the MRS to fall, i.e. the indifference curves to flatten (in each case, we plot only the indifference curve which passes through alternative A).

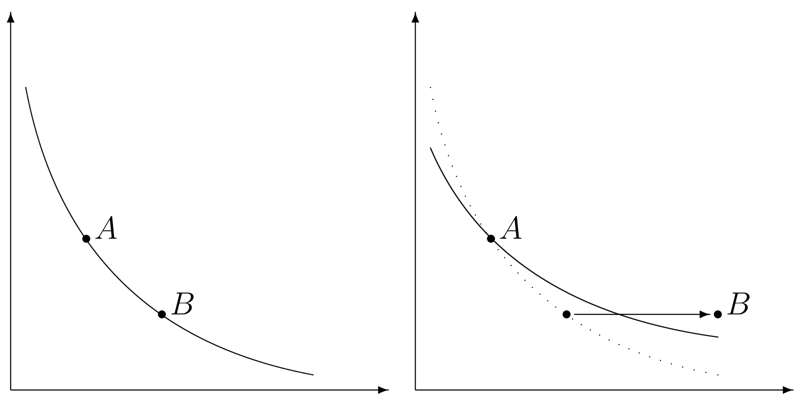


Figure 1: The Effect of Changes in the Choice Set on Indifference Curves

4 Relative Thinking in Market 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.

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 consumers has a vector of idiosyncratic valuations $v = (v_1, ..., v_N) \in \mathbb{R}^N$. 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:[4]

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

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.[5]

Each consumer will thus choose the 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

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

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

a constant addition to all prices, i.e. it is preserved when prices are transformed $p_i = p_i + \tau$.[6] 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,

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

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 we 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).[7] The model thus encompasses both horizontal and vertical differentiation.[8]

This allows our first comparative static,

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

This simply states that, because demand depends only on *differences* between prices, the equilibrium in relative prices will be independent of the level of cost, in other words aggregate pass-through will be exactly 1.

[6]This depends of course on $\underline{u}$ being sufficiently low.

[7]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 underly a demand system with an equilibrium.

[8]Typically in vertical differentiation models, consumers have the same preferred ranking among alternatives; in horizontal differentiation, they have different rankings. Because we 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.

4.1 Relative Thinking

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.[9]

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

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 relative thinking, as earlier defined.

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 we 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 2. *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.

[9] 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.

Proposition 3. *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 $\beta(p_j - p_i)$, and on nothing else. Thus, if all prices increase by the same amount, the first-order conditions will still hold. Likewise if β increases by some factor λ, 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.

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.

4.2 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 the effect on markups remains, at least in the simplified case of symmetric competition on a circle (a Salop model).

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

Proposition 4. *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)).

5 Robustness

5.1 Endogenising β

The model presented does not allow firms to take into account the effect of their own price on 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.

5.2 Income Effects

I have assumed that the underlying utility function is linearly separable in money, 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 seems to predict the opposite.

6 Notes on Interpretation

The predictions from relative thinking hinge on the contents of the choice set, and in many situations this may be unobservable.

This is a common problem with behavioural theories. Typical choice functions in economics depend only on objective outcomes, e.g. streams of consumption. It is often argued that choice also depends on subjective factors such as the level of consumption relative to a reference point; the source of income; the framing of a decision. These subjective variables are typically not observed in economic situations of interest, for example when predicting consumption from a tax rebate, the effect of a subsidy, or the decision to enroll in college, the subjective elements of the decision are not observed, so these theories are difficult to test.[10]

[10] Kőszegi and Rabin (2006) show how the reference point can be endogenised, i.e. made a subject of the choice set, in a reference-dependent theory.

In the theory presented here, the composition of the choice set is an unobservable subjective parameter.

In the most general sense, a person's choice set at each point is the set of the possible stochastic streams of consumption for the rest of their life. In this case, an increase in the price of a calculator should equally affect all their tradeoffs involving money, not just their choice of where to buy a calculator. So when this theory specifies a choice set, A, this should be thought of as a consideration set, meaning a set of salient alternatives at a given moment.[11]

I conjecture two common sets of salient alternatives. First, the same item, offered at different prices at different stores. This has been the principal focus of the paper.

An alternative choice set could be choosing between different varieties within a product category, side by side on a shelf. The analysis of pricing with multiple products is complicated, but in general relative thinking should introduce an incentive to raise prices, thus lowering β, and increasing demand at every price. If we allow the firm to introduce new products, this may also explain the decoy effect (sometimes called "compromise effect"), where introducing a high-price product shifts demand from low-price to medium-price options (Tversky and Simonson (1993), Krishna et al. (2006)).

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.

[11] The problem of too-large choice sets afflicts virtually any theory of menu-dependent preferences. There is some discussion of the difference between a choice set and a consideration set in Koszegi and Szeidl (2011).

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. 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.[12]

Of more interest is the oligopoly case, because all the studies we cite consider homogenous goods sold at different outlets. As we 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.[13] 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.

[12] 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.

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

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.

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 analyse 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 2 plots item cost against item markup, and shows the very strong relationship between the two variables. A $1.00 increase in cost 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 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 with cost above $10 has a markup below $1.[14]

[14] Huggies nappies and Pampers nappies.

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, we expect the positive predicted relationship.

Second, if firms sell multiple products then the optimal markup depends on interactions: if high-cost goods tend to be substitutes, and low-cost goods tend to be complements, this would again predict the observed relationship. A strong source of complementarity would come from any fixed cost of a visit to a store; 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 which is certainly higher for high-cost goods is simply the cost of capital). Second, there may be some goods which have high cost, and their optimal margin would be low, but which are not sold. For example, if they are 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 pricing, marking up by a constant fraction, even when it departs from the profit maximising price. Looking again at figure 2, 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.

7.2 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 we describe here are in the cross section, and so are subject to the caveats that we 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.[15]

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)

[15]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."

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.

8 Related Literature

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

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 modelled as transport costs increasing in price.[16]

This extends Azar's work in a number of ways: in deriving the behavioural

[16]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.

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.

9 Conclusion

This paper has shown how relative thinking will change a system of demand functions, which would ordinarily imply constant markups, such that in equilibrium higher cost will be associated with higher markups, higher dispersion, and more entry. We have shown that data supports the predictions, both in time series and in cross-section, though other explanations are also consistent with the data. Finally we introduced a large dataset of costs and prices from a drugstore, which shows a very tight relationship between cost and price.

References

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

Azar, O. (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**(5): 684–697.

Azar, O. (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**(8): 1077–1098.

Barzel, Y. (1976). An alternative approach to the analysis of taxation, *The Journal of Political Economy* **84**(6): 1177.

Baye, M., Morgan, J. and Scholten, P. (2004). Price dispersion in the small and in the large: Evidence from an internet price comparison site, *The Journal of Industrial Economics* **52**(4): 463–496.

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

Besley, T. and Rosen, H. (1999). Sales taxes and prices: an empirical analysis, *National Tax Journal* **52**(2): 157–178.

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

Clay, K., Krishnan, R., Wolff, E. and Fernandes, D. (2002). Retail strategies on the web: Price and non–price competition in the online book industry, *The Journal of Industrial Economics* **50**(3): 351–367.

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

Eichenbaum, M., Jaimovich, N. and Rebelo, S. (2011). Reference prices, costs, and nominal rigidities, *American Economic Review* **101**(1): 234–262.

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

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

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

Hoomissen, T. V. (1988). Price dispersion and inflation: Evidence from israel, *Journal of Political Economy* **96**(6): pp. 1303–1314.

Jaeger, D. and Storchmann, K. (2011). Wine retail price dispersion in the united states: Searching for expensive wines?, *American Economic Review* **101**(3): 136–41.

Kenkel, D. (2005). Are alcohol tax hikes fully passed through to prices? evidence from alaska, *The American economic review* **95**(2): 273–277.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences, *The Quarterly Journal of Economics* **121**(4): 1133.

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

Krishna, A., Wagner, M., Yoon, C. and Adaval, R. (2006). Effects of extreme-priced products on consumer reservation prices, *Journal of Consumer Psychology* **16**(2): 176.

Lach, S. (2002). Existence and persistence of price dispersion: an empirical analysis, *Review of Economics and Statistics* **84**(3): 433–444.

Pan, X., Ratchford, B. and Shankar, V. (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.

Poterba, J. (1996). Retail price reactions to changes in state and local sales taxes, *National Tax Journal* **49**(2): 165–176.

Pratt, J., Wise, D. and Zeckhauser, R. (1979). Price differences in almost competitive markets, *The Quarterly Journal of Economics* **93**(2): 189.

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

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

Sorensen, A. (2000). Equilibrium price dispersion in retail markets for prescription drugs, *Journal of Political Economy* **108**(4): 833–850.

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

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 Kahneman, D. (1981). The framing of decisions and the psychology of choice, *Science* **211**(4481): 453–458.

Tversky, A. and Simonson, I. (1993). Context-dependent preferences, *Management Science* **39**(10): 1179–1189.

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

Weyl, E. and Fabinger, M. (2009). Pass-through as an economic tool, Working Paper.

Young, D. and Bielinska-Kwapisz, A. (2002). Alcohol taxes and beverage prices, *National Tax Journal* **55**(1): 57–74.

Appendix 1: 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, stationary, 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 (28 March - 3 April, 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 have talked to staff at the store, one of whom confirmed that the code represents cost. One of the interesting finally, 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."[17]

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

I have asked two employees about the cost code. One was not aware of its meaning. The other knew the code. She said that they use it for making decisions about items to promote, especially when they have a lot of leftover stock. She also said that it was not used much, because most pricing decisions are made at the firm's headquarters.

Appendix 2: Proofs

Proof of Proposition 1. 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 2. 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.[18] 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. □

Proof of Proposition 3. 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 $p_j > p_i$. □

Proof of Proposition 4. 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

$$(p - c)\frac{1}{n} = C$$

[18] 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.

$$(p-c)\beta = \frac{1}{n}$$

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:

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

and

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

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

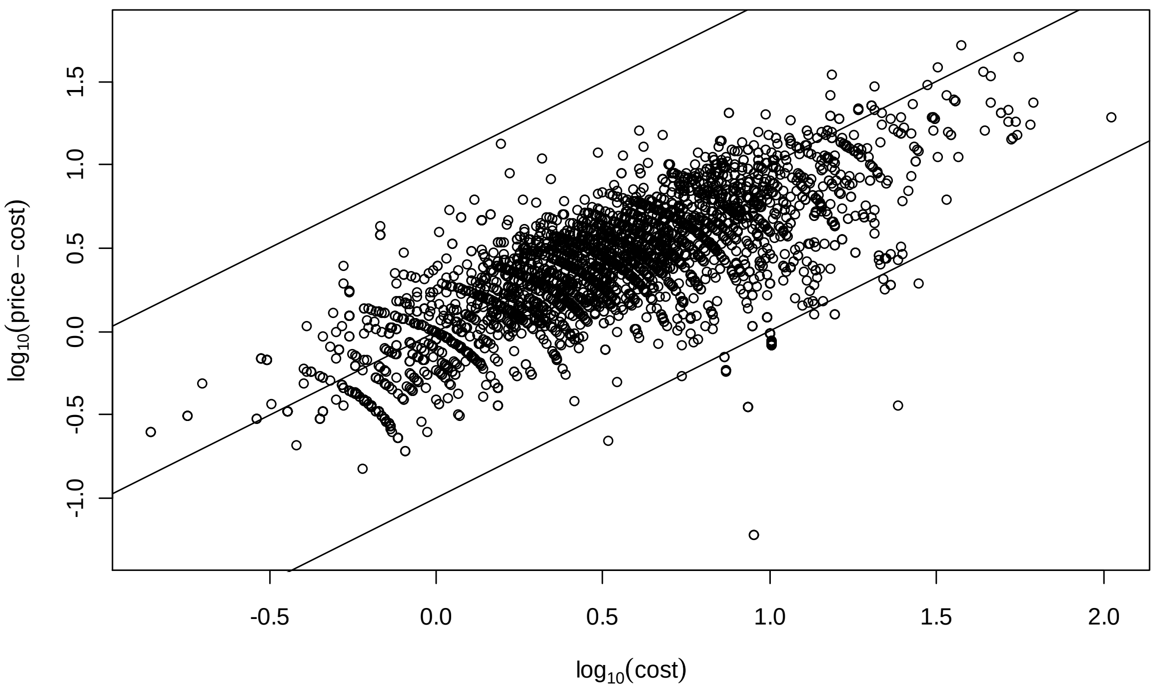


Figure 2: Cost and Markup in 3,500 Items from a Drugstore.
Both axes are plotted on a base-10 log scale; each of the three upward-sloping lines thus represent constant proportional markup rates of 10%, 100%, and 1000%. The curved pattens in the data points are caused by clustering of prices at common price-points: 99 cents, $1.99, etc.

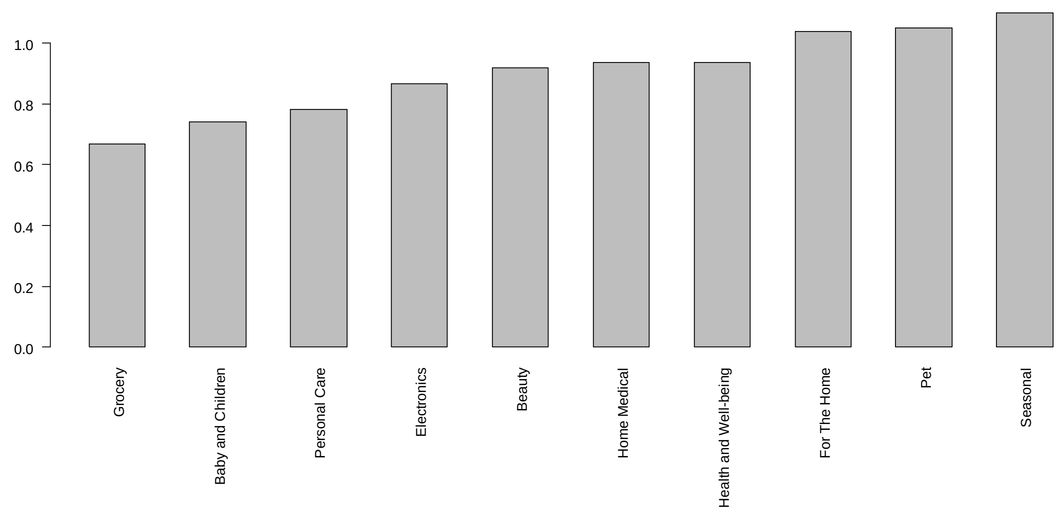


Figure 3: Average Proportional Markup ($\frac{p-c}{p}$) by Product Category