A Model of Reference-Dependent Preferences

Botond Kőszegi and Matthew Rabin

Department of Economics, University of California – Berkeley

March 2004

Acknowledgments: We thank Paige Skiba for research assistance, and Stefano DellaVigna, Mathias Dewatripont, Andy Postlewaite, Antonio Rangel, Jacob Sagi, Jean Tirole, seminar participants at Berkeley, the Harvard-MIT Theory Seminar, Rice, SITE, and UNH, and graduate students in courses at Berkeley and Harvard for helpful comments. Kőszegi thanks MIT for its hospitality while some of this work was completed, and Rabin thanks the Macarthur Foundation and the National Science Foundation for financial support.

Abstract

We develop a model that fleshes out, extends, and modifies existing models of reference-dependent preferences and loss aversion while accomodating most of the evidence motivating these models. Our approach makes reference-dependent theory more broadly applicable by avoiding some of the ways that prevailing models—if applied literally and without ancillary assumptions—make variously weak and incorrect predictions. Our model combines the reference-dependent gain-loss utility with standard economic "consumption utility" and clarifies the relationship between the two. Most importantly, we posit that a person's reference point is her recent expectations about outcomes (rather than the status quo), and assume that behavior accords to a *personal equilibrium*: The person maximizes utility given her rational expectations about outcomes, where these expectations depend on her own anticipated behavior. We apply our theory to consumer behavior, and emphasize that a consumer's willingness to pay for a good is endogenously determined by the market distribution of prices and how she expects to respond to these prices. Because a buyer's willingness to buy depends on whether she anticipates buying the good, for a range of market prices there are multiple personal equilibria. This multiplicity disappears when the consumer is sufficiently uncertain about the price she will face. Because paying more than she anticipated induces a sense of loss in the buyer, the lower the prices at which she expects to buy the lower will be her willingness to pay. In some situations, a known stochastic decrease in prices can even lower the quantity demanded.

Keywords: Endowment Effect, Hausdorff Metric, Loss Aversion, More Mugs, Mugs, Prospect Theory, Reference-Dependent Preferences

JEL Classification: B49

1 Introduction

How a person assesses the outcome of a choice is often determined as much by its contrast with a reference point as by an intrinsic preference for the outcome itself. In experimental and empirical investigations, the most noticeable manifestation of such reference-dependent preferences is loss aversion: Losses resonate more than same-sized gains. In purchasing decisions, the minimal acceptable selling price for an object is typically higher than the maximum price at which people are willing to buy the same object.[1] This "endowment effect" occurs partly because subjects construe selling the object as a loss, and consider buying it as (merely) a gain. And as emphasized in the first prominent formal model of reference-dependent preferences, Kahneman and Tversky (1979), the most significant source of risk aversion over modest stakes is aversion to losses.

It is becoming widely recognized that reference dependence and loss aversion may have important economic consequences, and researchers have begun to apply these ideas in a handful of economic situations.[2] Yet existing models are better suited to explaining experimental data, or to applying them in a specific context, rather than to systematically integrating them into economic theory. If applied literally and without ancillary assumptions, these models also make variously bad or weak predictions in many relevant potential applications. In this paper we build on the essential intuitions in Kahneman and Tversky (1979) and subsequent models of reference dependence, but flesh out, extend, and modify these models to build a realistic and more general theory of reference-dependent preferences that can be systematically applied to a wide array of economic settings. We demonstrate such applicability by establishing some strong predictions of the model in the analysis of consumer purchasing behavior.

We present the basic framework in Section 2. A person's utility depends not only on consumption, c, as in the standard neoclassical formulation, but also on a "reference" consumption level, r. Both c and r are K-dimensional vectors. To construct the utility function $u(c|r)$, we begin with an intrinsic "consumption utility" $m(c)$ that is independent of the reference level, and that can be interpreted as corresponding to the classical notion of outcome-based utility. Overall

[1] See Kahneman, Knetch and Thaler (1990, 1991).

[2] See, e.g., Benartzi and Thaler (1995) Barberis, Huang, and Santos (2001), and Genesove and Mayer (2001).

utility is given by $u(c|r) \equiv m(c) + n(c|r)$, where $n(c|r)$ is gain-loss utility. We assume both consumption utility and gain-loss utility are separable across dimensions, so that $m(c) \equiv \sum_k m_k(c_k)$ and $n(c|r) \equiv \sum_k n_k(c_k|r_k)$. The person's "gain-loss utility" in dimension k, $n_k(c_k|r_k)$, depends solely, through a "universal gain-loss function" μ, on how consumption utility in that dimension compares to what the person could have achieved with the reference consumption level: $n_k(c_k|r_k) \equiv \mu(m_k(c_k) - m_k(r_k))$. We posit that μ satisfies the properties of Kahneman and Tversky's (1979) value function, including a concave kink at zero that captures loss aversion.

While tying the two together so tightly is likely to lead to incorrect predictions in some situation, we feel that basing gain-loss utility on consumption utility is roughly correct—and is an important novel restriction for a reference-dependent model to make sensible and strong predictions in many economic contexts.[3] By tying gain-loss utility in different dimensions to the consumption utility in those dimensions, our model predicts for instance that people are less bothered by risk in goods of lower consumption value than by risk in dimensions of greater consumption value. And by adding consumption utility to the decisionmaker's utility function, it both replicates the predictions of Kahneman and Tversky's prospect theory value function under typical situations, where the consumption values of gains and losses are likely to be similar, and improves predictions in cases where the value function over consumption levels clearly does not apply. If a person's reference point in water is one quart below the survival level, for instance, her observed value function would surely not exhibit loss aversion: a one-quart increase in water endowment would be assessed as more valuable than a one-quart decrease would be assessed as detrimental.

Since behavior depends on the reference level, the predictions of reference-dependent theories in most settings depend crucially on what this reference level is assumed to be. Yet theoretical and empirical research on this issue is far less extensive than research on how people react to departures from a given reference point. A broadly realistic, but general and precise, theory of reference-point determination would seem useful. For a variety of reasons detailed in Section 3, we posit that a person's reference point is her recent beliefs about future outcomes. An employee who confidently

[3]Our formulation of this relationship relies on the assumption that consumption utility itself is separable across dimensions. While the model could be generalized, we argue in Section 5 that such separability is in fact conceptually compelling once consumption dimensions are appropriately chosen.

expects a 10% pay raise might assess a raise of only 5% as a loss. And if expecting Johnny Depp to star in a movie, it is a painful loss if he has merely a cameo.

While existing experimental evidence is generally interpreted in terms of the reference point being the endowment or status quo, we feel that virtually all of this evidence can also be interpreted in terms of expectations—for the simple reason that in the contexts studied people plausibly expect to keep the status quo. Among the array of situations classically studied by economists, however, there are many cases where people very much do not expect to keep the status quo. In most such cases, we believe equating the reference point with expectations will typically make better predictions. Eating meals and enjoying entertainment and many other activities are not part of a person's endowment prior to consumption; status-quo models are then seemingly forced to assign the same reference point of zero in all cases, so that irrespective of expectations all such consumption would be assessed as gains. And more fundamentally, most of the market activity that economists study is between buyers who *hope and expect* to lose money and gain items and sellers who *hope and expect* to gain money and lose items. While our analysis below focuses on some dramatic implications of loss aversion and reference dependence in market settings, it also implies that it is incorrect to extrapolate endowment-effect experiments by predicting market participants assess their planned trades in terms of losses. From the perspective of our model, when a consumer goes shopping, she does not assess carrying out planned purchases as involving a loss; rather, it is failures to carry out intended purchases, or paying more than expected, that are assessed as involving losses.

Saying that the reference point is expectations means that our predictions are driven in a large part by our theory of how expectations are formed. We complete our model by assuming rational expectations, and, along the lines of Kőszegi (2001, 2003), defining a "personal equilibrium" as a situation where the stochastic outcome implied by optimal behavior conditional on expectations coincides with expectations. While our approach takes analysis to its rational-expectations extreme, it provides a disciplined way of capturing the realistic assumption that people typically have some ability to predict their own behavior.[4]

[4] Our designation of expectations as the reference point makes another limitation of most previous models more apparent. Because an agent's beliefs are typically stochastic, applying reference-dependent utility theory must allow

In Section 4 we analyze consumer demand. Positing expectations as the reference point not only avoids the mis-extrapolation of endowment-effect findings to actual makets, it makes a series of strong predictions all centering around a central implication: A buyer's willingness to pay for a good does not reflect merely her intrinsic valuation for the good; rather, it also depends strongly on beliefs about whether she is going to buy it and how much she is going to pay for it. As such, our model predicts that consumer preferences may be significantly affected by market conditions. In this sense, it adds to other recent research suggesting that preferences cannot be taken as exogenous in studying market outcomes.[5] In addition, our model provides structure for investigating how and when the market conditions affect preferences.

One striking implication of the endogenous determination of consumer valuation is the possibility of multiple personal equilibria. Intuitively, if a customer expects to buy a pair of shoes, she construes coming home without a pair as a (strongly felt) loss of shoes not acquired and a (less important) gain of money not paid, making her inclined towards buying. Conversely, if she expects not to buy, she experiences giving up money as a loss, and getting shoes as merely a gain, making her disinclined to buy. For a range of market price levels, this generates multiple equilibria.

A more interesting implication of the endogenous determination of consumer valuation is that the buyer's reservation price depends on the distribution of prices she is facing. An extreme manifestation of this dependence is that a known stochastic downward shift in prices can *lower* a consumer's demand for a good. There are always prices p_H and $p_L < p_H$, for example, such that if a consumer knows that the price is p_H, it is an equilibrium for her to buy for sure; but if she assigns equal probability to each of the two prices, the unique equilibrium is never to buy. Intuitively, anticipation of the possibility of buying at a low price makes paying the high one feel

for risky reference points. In our model, therefore, utility depends both on the probability distribution over consumption levels $F(c)$ and a reference point $G(r)$ that is itself a probability distribution. We assume that in comparing a given outcome c to the reference point G, the agent compares c to all r possible under G, and takes the average of these gain and loss sensations.

[5]See, e.g., Ariely, Loewenstein, and Prelec (2003), on how preferences that appear coherent and consistent may in fact be relatively arbitrary, and Gul and Pesendorfer (2004), who discuss some implications for consumption behavior of assuming that preferences depend on choice sets. Like our model, these theories say that the set of opportunities induced by markets will influence preferences.

like a loss. This "comparison effect" can make a consumer unwilling to buy at the high price. But if she anticipates not buying at the high price, she becomes less attached to the idea of buying, and does not buy at the low price, either.[6] More generally, since high prices are always subject to an unfavorable comparison with lower prices, lowering prices at which the consumer would otherwise have bought exerts a downward pressure on her reservation price.

Yet the above kind of probabilistic "sale" can also *increase* demand. There are prices p'_H and $p'_L < p'_H$ such that if the price is p'_H with certainty, it is an equilibrium for the consumer not to buy, but if the price is uncertain, it is the unique equilibrium for her to buy at both prices. Intuitively, p'_L is so low that the consumer would buy at that price in any equilibrium. But once she anticipates buying with some probability, she feels more of a loss if she does not buy. This "attachment effect," a possible motive behind some promotions and sales techniques, induces her to buy at price p'_H as well. Section 4 generalizes these effects on the consumer's reservation price, and derives a number of other properties of personal equilibrium in consumer markets.

For most of the paper we follow the common interpretation in the literature that people assess gains and losses separately in each of the goods being considered. As we argue in Section 5, however, a more accurate view is that people assess gains and losses in dimensions of hedonic experience that may not correspond to the set of physical goods whose levels change, and the assessment of gains and losses when trading off different goods depends on how similar they are hedonically. For example, a restaurant goer who faces a set of forty possible dishes might psychologically categorize them by type (not individual dish), so that she experiences no loss if she eats any one of the twenty Chinese dishes she thought possible, somewhat of a loss if she eats an Italian meal—and a significant loss if she goes to the movies instead of dinner. More speculatively, it may also be that goods that are hedonically distinct come to be integrated by consumers who become accustomed to substituting them. If a diner becomes accustomed to randomly eating either of two gastronomically dissimilar dishes, she may come to assess the consumption of neither as a loss. Accommodating such possibilities requires substantive assumptions from outside the model and changes the predictions,

[6]As this intuition suggests, the violation of the law of demand occurs due to the equilibrium feedback of the market price and behavior into expectations. For any *given* expectations, demand curves are downward-sloping.

but we show in Section 5 that our model can fully accommodate these phenomena by applying it to the appropriate "hedonic" dimensions.

In addition to the ways we believe our model substantially clarifies and improves on existing theories of reference-dependent preferences, it also has an attractive "methodological" feature in promoting the addition of reference dependence to existing economic models. By deriving gain-loss utility in a fixed way from consumption utility and by fully endogenizing the reference point as rational expectations of the outcome given a specific economic situation, our theory brings us closer to a universally applicable way to translate any existing reference-independent model into the corresponding reference-dependent one. Ideally, once the universal gain-loss function, μ, is specified, applying our analysis to a given situation would require exactly the same assumptions that economists using the classical model apply.

The model in this paper does not accomplish this ambitious goal, however. While we believe applying the model is straightforward in most cases, the necessity of using some psychological and economic judgment—for instance, in choosing the appropriate hedonic dimensions and the appropriate notion of "recent expectations"—leaves us significantly short of an entirely formulaic way to extend the classical utility model. Several other shortcomings of our analysis are clear. Many of our specific assumptions are based on intuition rather than direct evidence. And there are also clearly some settings—most importantly, where a person's reference point acclimates to the consequences of a choice long before experiencing those consequences—where the same principles motivating our approach would in fact lead to a different reduced-form model. We discuss such shortcomings and gaps, some possible resolutions, as well as further economic applications of our model, in Section 6.

2 Reference-Dependent Utility

To develop our basic model, we start with riskless outcomes. The person's utility is given by $u(c|r)$, where $c = (c_1, c_2, \ldots, c_K) \in \mathbb{R}^K$ is consumption and $r = (r_1, r_2, ..., r_K) \in \mathbb{R}^K$ is a "reference level" of consumption. For now it is easiest to think of the vector c as a vector of consumption goods.

To capture preferences over risky outcomes, suppose that c is drawn according to the probability measure F. Then the person's utility $U(F|r)$ is given by

$$U(F|r) = \int u(c|r)dF(c). \tag{1}$$

Our model extends reference-dependent preferences to situations in which the reference point itself might also be a probability measure. Such an allowance is clearly necessary in our framework, where we assume the reference point is beliefs about outcomes. Suppose the person's reference point is the probability measure G over $\mathbb{R}^K$, and her consumption is drawn according to the probability measure F. Then, her utility is

$$U(F|G) = \int_c \int_r u(c|r)dF(c)dG(r). \tag{2}$$

Our formulation of reference-dependent utility captures, in a simple way, the notion that the sense of gain or loss from a given consumption outcome derives from comparing it to all outcomes in the support of the reference lottery. For example, if the reference lottery is a gamble between \$0 and \$100, an outcome of \$0 feels neutral relative to \$0, and like a loss relative to \$100, and the overall sensation of loss increases with the probability of \$100 in the reference lottery. That a person's utility depends on a reference lottery in addition to the actual outcome is similar to Gul's (1991) model of "disappointment aversion," although his theory of behavior is very different from the one we develop in Section 3.[7] Our utility function is also closely related to Sugden's (2003) recent model of reference-dependent preferences.[8] Sugden, however, does not provide a theory of reference point determination, as we do below.[9]

[7]Also, Gul's formulation in effect amounts to replacing the reference consumption lottery with its certainty equivalent (according to a reference-independent utility function), to which the utility from actual consumption is then compared. Thus, two reference lotteries that have the same expected utility generate the same behavior. As a result, our utility function has similar properties to Gul's when the reference point is riskless, but not when it is risky.

[8]The main difference between Sugden's utility function and ours is in the way a given consumption outcome is compared to the reference lottery. In our model, each outcome is compared to all outcomes in the support of the reference lottery. In Sugden (2003), an outcome is compared only to the outcome that would have resulted from the reference lottery *in the same state.*

[9]Due to its descriptive nature, our theory bears less resemblance to two recent models of reference dependence and

Contrary to the models of Kahneman and Tversky (1979), Machina (1982), and subsequent models building on these theories, we assume that a person's preferences are linear in probabilities given the reference point. The evidence is clear that people's evaluation of prospects is not linear in probabilities. We omit this feature of preferences to keep our model simple, doing so with the impression that there are few interesting interaction effects between such non-linearities and the central features of our model.

While reference levels matter a lot, the gain or loss relative to a reference point is not *all* that people care about. While people assess losing a mug as an unpleasant experience, they may also care about having the use of a mug the next day. And clearly some of the benefits of having more money is not just the sensation of having and using more money—but rather the absolute pleasure of consumption from the goods we spend it on. Hence, we assume that overall utility has two components: $u(c|r) \equiv m(c) + n(c|r)$, where $m(c)$ is "consumption utility" classically stressed in economics, and $n(c|r)$ is "gain-loss utility." This decomposition is of course mathematically meaningless without the attending assumptions we make about the functions $m(c)$ and $n(c|r)$. We now turn to specifying and motivating the assumptions we make about these two functions. Among other advantages, we argue below that our approach of explicitly incorporating consumption utility into the analysis is clearly more complete in terms of both behavior and welfare than formulations with a single "value function" that evaluates gains and losses relative to a reference point, and ignores or suppresses the role of consumption utility in the evaluation of outcomes. Perhaps more importantly, our approach below facilitates the "translation" of a classical (reference-independent) model into a reference-dependent one by taking the utility function from that model and identifying it with our $m(\cdot)$.

status quo bias, Masatlioglu and Ok (2002) and Sagi (2002). These authors start by formulating normative axioms that allow for reference effects, and derive the utility representations these axioms imply. Our focus, instead, is on building a descriptive model that is broadly consistent with a variety of evidence. As a result, our theory does not satisfy some of the basic normative postulates these authors propose, such as both papers' "no-cycling" assumption and the feature of Masatlioglu and Ok's theory that the status quo is irrelevant for choice once it is abandoned. (And, while we specifically do *not* think of our model as normative, we also do not find these particular axioms normatively compelling.)

Throughout the paper, we restrict attention to consumption-utility functions m which are additively separable across dimensions: $m(c) \equiv \sum_{k=1}^{K} m_k(c_k)$, where the functions m_k are differentiable and strictly increasing. At the cost of some complication, the framework could be generalized to nonseparable consumption utility. Provocatively, however, we will argue in Section 5 that for the purposes of specifying reference-dependent preferences it is always appropriate to reformulate utility theory into hedonic dimensions that are separable.

We assume that a person assesses gain-loss utility in each dimension separately: $n(c|r) \equiv \sum_{k=1}^{K} n_k(c_k|r_k)$. Thus, in evaluating how satisfied she is with the outcome, the decisionmaker "breaks up" deviations in consumption from the reference level into deviations in individual dimensions. In combination with loss aversion, this separability of gain-loss utility across dimensions is at the crux of many implications of reference-dependent utility, including the endowment effect.

Beyond saying that a person cares about consumption in addition to gain-loss utility, we also propose that there is a strong relationship between the two. Consider the following thought experiment. A person is endowed with 100 paper clips and 100 \$10 bills as part of her reference point, and must choose between two gambles: a 50-50 chance of gaining a paper clip or losing a paper clip, and the comparable gamble involving \$10 bills. It seems likely that she would risk losing the paper clip rather than the money, and do so because her sensation of gains and losses is generally likely to be smaller for a good whose consumption utility is smaller: Insofar as her concern for a paper clip is negligible relative to her concern for \$10, the paper-clip lottery is essentially a way to reject the monetary gamble.[10] Yet if m is approximately linear for such small stakes, the choice depends almost entirely on n, so that the prediction must derive from a comparison of the relative slopes of gain-loss utility in the paper-clip dimension and gain-loss utility in the money dimension. Any model of reference-dependent preferences that does not relate gain-loss assessments to consumption utility is not equipped to provide guidance in this or related examples.

While it surely exaggerates the tight connection between the two components, our model assumes that how a person feels about gaining or losing in a dimension depends solely on the changes

[10]In hypothetical questions to our classes totalling over eighty students, virtually everyone chose to risk losing the paper clip. We have not run the experiment with real paper clips.

in consumption utility associated with such gains or losses. Specifically, we assume that:

$$n_k(c_k|r_k) \equiv \mu\left(m_k(c_k) - m_k(r_k)\right), \tag{3}$$

where $\mu(\cdot)$ is a "universal gain-loss function." This says that the evaluation of gain or loss in dimension k depends in a universal way on the change in a person's dimension-k consumption utility caused by the deviation from the reference point.

Inspired by the model in Kahneman and Tversky (1979), as enhanced by Bowman, Minehart, and Rabin (1999), we assume that μ satisfies the following properties:

A0. $\mu(x)$ is continuous for all x, twice differentiable for $x \neq 0$, and $\mu(0) = 0$.

A1. $\mu(x)$ is strictly increasing.

A2. If $y > x > 0$, then $\mu(y) + \mu(-y) < \mu(x) + \mu(-x)$.

A3. $\mu''(x) \leq 0$ for $x > 0$ and $\mu''(x) \geq 0$ for $x < 0$.

A4. $\frac{\mu'_+(0)}{\mu'_-(0)} \equiv \lambda > 1$, where $\mu'_+(0) \equiv \lim_{x\to 0} \mu'(|x|)$ and $\mu'_-(0) \equiv \lim_{x\to 0} \mu'(-|x|)$.

Assumptions A2 and A4 capture the key property of reference-dependent preferences, that losses resonate more strongly than gains. A2 captures this notion of loss aversion for "large" stakes, and A4 for "small" ones. The slope of μ determines how much the person cares about departures from her reference point relative to consumption utility: fixing λ, the importance of gain-loss utility is increasing in $\mu'_+(0)$.

Assumption A3 captures another important feature of how people assess departures from their reference levels. Namely, preferences exhibit diminishing sensitivity: The marginal change in gain-loss sensations is greater for changes that are close to one's reference level than for changes that are further away. While the inequalities in A3 are most realistically considered strict to capture diminishing sensitivity, we shall often be interested in characterizing the implications of reference dependence where diminishing sensitivity does not play a big role. For doing so, we define an alternative to A3 that isolates loss aversion in our model by eliminating the diminishing sensitivity.

A3′. For all $x \neq 0, \mu''(x) = 0$.

Our first observation about the model is that it captures two basic aspects commonly associated with reference-dependent preferences. First, that fixing the outcome, a person is happier with a lower reference point than with a higher one (Part 1); and second, that preferences exhibit a status quo bias (Parts 2 and 3):

Proposition 1 *If μ satisfies Assumptions A0-A4, then the following hold.*

1. *For all F, G, G' such that for all $k \in \{1, \ldots, K\}$, the marginal G'_k first-order stochastically dominates G_k, $U(F|G) \geq U(F|G')$.*

2. *For any $c, c' \in \mathbb{R}^K$, $c \neq c', u(c|c') \geq u(c'|c') \Rightarrow u(c|c) > u(c'|c)$.*

3. *Suppose μ satisfies A3′. Then, for any F, F' such that $F \neq F'$, $U(F|F') \geq U(F'|F') \Rightarrow U(F|F) > U(F'|F)$.*

Parts 2 and 3 of Proposition 1 mean that if a person is willing to abandon her reference point for an alternative, then she prefers the alternative if that is her reference point. Under Assumptions A0-A4, this is always true for riskless consumption bundles, and is an immediate consequence of loss aversion. For a bundle c to be preferred to the reference point c' in a situation where c' is better along some dimensions, c must offer larger gains over c' in the dimensions in which it is better. And in that case, to avoid large losses in these dimensions, the person would not choose c' when the reference point is c. But the analogous statement for lotteries over consumption bundles requires the more restrictive assumption A3′. If a person has little diminishing sensitivity in the loss domain, and substantial diminishing sensitivity in the gain domain, she would rather suffer small gains and losses than large ones. Thus, she might prefer F over the reference point F' because it offers a number of small gains and losses instead of fewer large gains and losses. But when the reference point is F, it may no longer be true that F offers enough small gains to be preferred to F'.[11] That the person likes to "stick to" her reference point is a crucial property in generating a

[11] For a counterexample, suppose that there are two dimensions, money and a consumption good. Let F be a lottery that gives each of the consumption vectors $(0,0), (10,0)$, and $(20,0)$ with equal probability, and let F' be a lottery

number of our results, including the endowment effect and multiple personal equilibria in Section 4 below.

In fact, as long as income effects are negligible, Part 2 of Proposition 1 implies the endowment effect in a straightforward way. Let $K = 2$, the two dimensions being "mugs" and "money," and normalize the person's current wealth level and endowment of other mugs to zero. Let $c' = (0, 0)$, and suppose p_b is the price for which $c = (1, -p_b)$ satisfies $u(c|c') = u(c'|c')$. That is, p_b is the person's "buying price:" it is the maximum amount of money for which she is still willing to buy the mug, when her endowment and reference point is not to buy it.[12] Then, by Proposition 1, $u(c|c) > u(c'|c)$. This means that if a person is endowed with a mug, she is not willing to trade it for p_b; equivalently, her "selling price" is greater than her buying price.

Although our specification of preferences is different from previous formulations, it shares some of their key properties. In fact, in an important special case, when m is linear, our utility function has exactly the same qualitative properties as Kahneman and Tversky's prospect theory modified to assume that decisions weights are linearized to be probabilities.

Proposition 2 *If m is linear and μ satisfies Assumptions A0-A4, then there exists $\{v_k\}_{k=1}^K$ satisfying Assumptions A0-A4 such that, for all c and r,*

$$u(c|r) - u(r|r) = \sum_{k=1}^{K} v_k(c_k - r_k). \tag{4}$$

that gives each of $(0, \frac{21}{72})$ and $(20, \frac{21}{72})$, with equal probability. F and F' each give the same expected money outcome, with F' riskier than F, but F' gives a deterministically better consumption outcome. Suppose that consumption utility is given by $m_1(x) = m_2(x) = x$, and that the universal gain-loss function is $\mu(x) = x$ for $x \leq 0$, $\mu(x) = 0.9x$ for $0 \leq x \leq 10$, and $\mu(x) = 9 + 0.05(x - 10)$ for $x \geq 10$. That is, the decisionmaker has no diminishing sensitivity in losses, but her sensitivity to gains drops substantially at $x = 10$. One can verify that $U(F|F) - U(F'|F) < 0$ and $U(F|F') - U(F'|F') > 0$. Intuitively, F is better than F' in the money dimension, because—irrespective of whether the agent's reference point is F or F'—it breaks up large gains and losses into smaller ones. However, since a reference point of F' generates larger gains and losses overall, this advantage is greater when the reference point is F'. Given that F' is slightly better than F in the consumption dimension, F is preferred when F' is the reference point, but not when F is the reference point.

[12]It is not obvious how our assumption that the reference point is the agent's endowment can be made consistent with our view that expectations are the reference point. We discuss this issue in Section 3.

Furthermore, for each k,

$$\lim_{x\searrow 0}\frac{v_k'(-x)}{v_k'(x)}=\frac{1+\mu_-'(0)}{1+\mu_+'(0)}\in(0,\lambda). \tag{5}$$

This special case is all the more important because for most small departures from the reference point, m can be taken to be approximately linear, at least relative to the curvature of μ. Thus, for small changes in case of a non-linear m as well as a linear consumption utility, our utility function shares the qualitative properties of standard formulations of prospect theory.

Of course, this equivalence holds only when we assume that the gains and losses are small enough not to significantly change the marginal utility of consumption. When the changes are larger, or when marginal utilities change very quickly in the dimension being examined, the equivalence breaks down. In this range, the properties we posit for μ which are inspired by what prospect theory posits for the observable "value function" do not necessarily hold for this observable value function.

This may, however, be a good thing. We believe it likely that when consumption changes lead to dramatic changes in the marginal consumption value of the goods, $u(c|r)-u(r|r)$ will violate Assumptions A0-A4 in the way predicted by our model. Suppose, for instance, that a person's endowment of water is two quarts above the consumption level needed for survival. Then, our model predicts that, if forced to choose one or the other, she would absorb a sure loss of one quart in water consumption rather than risk a one-third chance of losing three quarts. Thus, her observable value function does not feature risk lovingness in losses. This reflects both the fact that her consumption utility is very concave in that region, *and* that her sense of loss from a one-quart drop in consumption is significantly smaller than her sense of loss from an additional one-quart decrease.[13] If, on the other hand, the person is just below the minimum sustainable water

[13]Similarly, when large losses in consumption or wealth are involved, we believe that diminishing marginal utility of wealth as economists conventionally conceive of it is likely to do battle with the diminishing sensitivity in losses emphasized in prospect theory. This perspective may help explain why the diminishing sensitivity in losses is less firmly established than the other features of the prospect-theory value function, especially when large losses are involved. It is our impression that both the ambiguity of features of the value function and the tension between

consumption—so that her consumption utility is very convex around her endowment—she would be much happier about a one-quart increase in water consumption than she would be unhappy about a one-quart decrease. Thus, our theory predicts that in this situation her observable value function does not feature loss aversion.

3 The Reference Point as (Endogenous) Expectations

Most laboratory studies of reference-dependent preferences analyze subjects' preferences over gains and losses in situations where the reference point is arguably clear—and usually the status quo. Unfortunately, while a mass of evidence regarding preferences *given* the reference level has been accumulated, there is much less empirical or theoretical research on the determinants of reference points.[14] Despite the shortage of evidence, for several reasons we take the strong view in our model that the reference point is determined by a person's expectations about what she is going to get.

Some evidence does indicate that expectations are more important than the status quo or lagged consumption in determining a person's sensation of gain or loss. In Mellers, Schwartz, and Ritov (1999), subjects' reported emotions after the outcome of a lottery depended systematically on both the outcome and the lottery's unobtained outcome. For example, subjects were (unsurprisingly) always happier with an $8 gain than with an $8 loss, holding constant the alternative possible outcome. But whichever outcome they obtained, they were happier if the alternative was a loss of $32 than if the alternative was a gain of $32. Although it is not clear whether brain imaging studies can be interpreted in terms of preferences, a suggestive finding by Breiter, Aharon, Kahneman, Dale, and Shizgal (2001) is that neural responses to the outcome of a lottery also depend on both

consumption utility and gain-loss utility have been intuited by prior researchers, but we are unaware of any model that addresses the issue.

[14]A passage in Tversky and Kahneman (1991, pp. 1046-47) is representative: "A treatment of reference-dependent choice raises two questions: what is the reference state, and how does it affect preferences? The present analysis focuses on the second question. We assume that the decision maker has a definite reference state X, and we investigate its impact on the choice between options. The question of the origin and the determinants of the reference state lies beyond the scope of the present article. Although the reference state usually corresponds to the decision maker's current position, it can also be influenced by aspirations, expectations, norms, and social comparisons."

the outcome and unobtained outcomes. More precisely, if a subject received \$0 in a lottery with possible outcomes of \$0, \$2.50, and \$10, activation in the sublenticular extended amygdala and nucleus accumbens regions of the brain was more negative than if she received \$0 in a lottery with possible outcomes of \$0, -\$1.50, and -\$6.

There is also evidence that expectations influence preferences through physiological changes. Certain types of consumption induce responses aimed at keeping our biological system in balance. But not only does one's body react to consumption that has already occurred, it can "prepare" for it beforehand. The expectation to consume is an important mediator in this *feedforward homeostatic regulation*, which changes the marginal utility of consumption.[15] Ever since Pavlov's dogs invented salivation, for instance, the expectation that a nice meal is imminent leads all of us to release saliva and stomach acids, making us hungrier.

Our hypothesis that the reference point is expectations may seem to fly in the face of most existing empirical research on loss aversion, which virtually all researchers have assumed either explicitly or implicitly to indicate that the reference level is the pre-choice status quo.[16] Yet we believe these same experiments can also naturally be interpreted in terms of expectations. Consider the classic endowment-effect experiment. If a subject is given a mug and is told it is hers to keep, her expected endowment of mugs ten minutes from now might reasonably be one mug. Thus, relative to expectations, not being able to keep the mug is a loss. In fact, in the typical experiment *all* subjects are given a mug (the buyers for inspection), but only some—the owners—are told it is theirs to keep. Arguably, the difference between owners and non-owners is not current or lagged physical possession, but rather expectation of future possession.[17]

[15] See Frederick and Loewenstein (1999) for a discussion of some of these issues.

[16] More generally, many researchers have posited that the reference point is a weighted sum of past consumption. (See Frederick and Loewenstein (1999) for a more detailed discussion of some of the specific functional forms that have been discussed.) Such an account of how reference levels are determined seems intuitive, but very little evidence has been gathered on its validity. One exception to this is Strahilevitz and Loewenstein (1998), who show that an agent's selling price for a mug (what else) increases with the time—in their experiment, up to 20 minutes—the mug has been in her endowment.

[17] In fact, some striking results by Plott and Zeiler (2003) can be interpreted as suggesting that if owners and non-owners do not differ in their expectation of future possession, the endowment effect is not observed. In their

Thought experiments on the implications of loss aversion in labor markets and in consumer behavior also suggest a central role for expectations. An employee who gets a 5% raise after she was expecting a 10% one probably does not experience much of a sensation of gain, in contrast to what reference-dependent preferences based on the status quo or lagged consumption would predict. Similarly, a consumer who goes to the store with the conscious expectation to buy a TV may not experience much of a loss from spending money on the TV. In fact, if a much-anticipated TV is not available, she will assess its lack as a loss. Thus, she experiences a loss only if she does *not* spend money.

For many experiences, a status-quo theory of the reference point would not only make wrong predictions, it would also have little substance. For example, if a person expects to undergo a painful dental procedure, finding out that she does not have to undergo it after all may feel like a gain. Yet she has not gained a "non-procedure" any more than somebody who never expected the procedure. In fact, there is no meaningful way in which different people have different status quo endowments of dental procedures, so irrespective of expectations a status quo theory would always predict the same gain-loss utility from these experiences. Since so many examples of reference dependence—including food, entertainment, and travel—involve fleeting consumption opportunities, where no good is in one's endowment prior to consumption, cases where the reference point cannot be the status quo may even be the typical environment where these models apply.

While we believe that expectations play a central role in determining reference points, we do not claim that they determine them completely. A clear exception to the heavy influence of expectations is sensory experiences. A person feels the contrast effects of walking out of a warm apartment into

experiment, subjects went through 2 unpaid and 14 paid practice rounds buying and selling lotteries. In the 15th paid round, then, Plott and Zeiler ran a classic endowment effect experiment using mugs, and observed no statistically significant difference between buying and selling prices—with point estimates actually contrary to the endowment effect. An interpretation of the reported results is that the practice rounds acclimate subjects to the prospect that they might soon give up goods they receive, and buy goods they do not receive. Under these circumstances, our theory predicts a substantial narrowing of the gap between buying and selling prices. (Our model would not, however, explain their empirical finding of a complete elimination of the effect.) See also List (2003), who finds that experienced sports card traders do not show an endowment effect. Once again, one interpretation of this finding is that more experienced traders come not to expect to necessarily keep any particular item they have just acquired.

the illegitimately cold Boston winter no matter how much she expects the cold.[18] While our model is unrealistically extreme in designating expectations as the sole reference point, we hope that, at a minimum, our concrete formulation will prompt experiments and other empirical work that directly study the nature of reference points.[19]

The above evidence and intuitions lead us to propose expectations as an appropriate theory of reference points. But doing so confronts us with some important conceptual issues. Most economic situations of interest involve choice, and what a person wants to do depends on her preferences. Therefore, so long as she has some ability to predict her own behavior, not only do preferences depend on expectations—expectations depend on preferences as well. If a person's expectations are not wholly exogenous and typically incorrect, this feedback must be taken into account. This suggests that it is appropriate to model even individual decisionmaking in terms of an "equilibrium" concept that requires a person's expectations to be consistent with her own eventual outcomes—which depend on both the choice set she happens to face and the choice she makes from that set.

In our formal model, the decisionmaker makes a single choice from some choice set $D \subset \Delta(\mathbb{R}^K)$.

[18]There is also evidence that we compare our outcomes to our aspirations and goals. Locke and Latham (1991) report that setting higher goals leads subjects to produce higher performance. And according to Heath, Larrick, and Wu (1999), a person who has failed to reach a goal feels worse than one who performed identically but did reach her goal. (Even here, however, a crisper notion of how goals influence expectations would prove essential to interpreting results.) In addition, it seems that people's experiences with outcomes are affected by relevant social comparisons. Clark and Oswald (1996) conclude that workers are less satisfied with their jobs if others with similar labor market characteristics have higher income. More generally, the literature on social comparison theory in psychology highlights that people tend to compare themselves to relevant others along a number of dimensions, with positive comparisons being pleasant, and negative comparisons being unpleasant.

[19]Experiments in which subjects' expectations are carefully controlled would prove especially fruitful. Ariely and Kőszegi have (in a never-to-be-published experimental study) attempted to do this. Half of the students at an MIT undergraduate class were told that they would receive a mug with a 90% probability, and half were told that they would receive it with 10% probability. Selling prices conditional on getting the mug were then elicited. The difference between the two conditions was statistically insignificant (in the direction predicted by our theory). While this result is inconsistent with our model, we feel that the notion of specifying the reference point as expectations is sufficiently compelling to warrant careful further study.

Before her decision, however, she may not know the choice set she is going to face. Rather, she may have probabilistic beliefs. Suppose that the collection of possible choice sets, $\{D_l\}_{l\in\mathbb{R}}$ is indexed continuously (in the Hausdorff metric) by the real numbers, and is determined probabilistically by the distribution Q over $\mathbb{R}$. Based on this, we define:

Definition 1 *A selection* $\{F_l \in D_l\}_{l\in\mathbb{R}}$ *is a* <u>*personal equilibrium*</u> *if for all* $l \in \mathbb{R}$ *and* $F'_l \in D_l$, $U\left(F_l \,|\, \int F_l dQ(l)\right) \geq U\left(F'_l \,|\, \int F_l dQ(l)\right)$.

If the person expects to choose F_l from choice set D_l, then (given her expectations over possible choice sets) she anticipates receiving the distribution of outcomes $\int F_l dQ(l)$. Definition 1 says that if this is the person's reference point, she should indeed be willing to choose F_l from the set D_l for each l.

The personal-equilibrium solution concept can be illustrated with the example of consumer purchases, which we will analyze in more detail in Section 4. Suppose a consumer goes to the store to possibly buy a pair shoes, but does not know their price until she gets to the store. Her choice set (over shoe-money allocations) at the time of the purchase decision depends of course on the unique realized price she ends up facing. The person's planned behavior in each choice set (i.e. her decision whether to buy, as a function of the price), combined with the distribution over possible choice sets (prices), implies a distribution of outcomes that becomes the decisionmaker's reference point. Personal equilibrium requires her planned behavior in each choice set to be optimal given this reference point.

Because this is simply an application to our environment of personal equilibrium as first defined in Kőszegi (2001,2003), Theorem 1 of Kőszegi (2003) establishes that, if each D_l is convex and compact, and they have a uniformly bounded support, a personal equilibrium exists.

The key idea behind our formulation is that a person's preferences depend on expectations she held after the time she started focusing on the problem, but prior to the time of consumption. That is, preferences depend on *lagged* expectations. If preferences depended solely on expectations after all decisions are made and all uncertainty is realized, our model would reduce to a solely outcome-based utility. But we posit that preferences do not react so quickly: while expectations about whether shoes are available may adjust immediately once the consumer enters the store, this adjustment in expectations does not immediately eliminate the consumer's comparison of the outcome to what she expected just prior to entering the store.[20] It bears emphasizing that our model is not based on the premise that beliefs are slow to adjust to new information—but that preferences do not immediately change when beliefs are revised.[21]

The other part of our psychological hypothesis is that preferences depend on expectations after the decisionmaker started focusing on the decision. Whether or not the classical model says people have beliefs about everything at all times, we assume such beliefs if they exist do not influence preferences until they come somewhat into focus. The specification of Q should incorporate this aspect of the person's decision problem, and is therefore an important interpretational matter in any application. As an illustration, consider again the shoe-shopping consumer who does not know the price of a pair of shoes until she gets to the store. Even in this fixed economic environment, Q depends on when the person started thinking about the decision. If she had been thinking about her possible purchase for a long time, her expectations from before she knew the price significantly affect her preferences. In this case, Q is the non-deterministic lottery representing her probabilistic beliefs over prices. It could also be that she only considered the possible purchase once she got to the store and found out the price. In this case, Q should be specified as the deterministic lottery corresponding to this price.

This setup also provides a way to interpret the situation subjects face in endowment-effect experiments, where we are inclined to believe behavior may typically be "non-equilibrium" in that subjects' expectations do not accord to what we researchers know to be the typical outcomes of experiments.[22] At the beginning of a typical experiment, subjects learn their role as either an

[20]Similarly, a drug addict who was expecting to shoot up, but now finds the drug unavailable, does not instantaneously stop craving the drug.

[21]Our notion of personal equilibrium is related to Geanakoplos, Pearce, and Stacchetti's (1989) notion of psychological Nash equilibrium in games. In both models, utility depends on beliefs. But in their case, these are higher-order beliefs about strategies, and in our case, they are about the agent's own outcomes. Another conceptual difference is our emphasis that preferences depend on *lagged* expectations. However, we have not fully explored the relationship between the two solution concepts.

[22]Although we demonstrate in Section 4 that endowment-effect type behavior would also result in personal equilibrium, probably few subjects reach such equilibrium within the time course of the experiment.

"owner" of a mug or a "non-owner". Clearly, subjects do not think of their mug possessions until this point. Plausibly, their reference point then becomes aligned with their ownership status. Once subjects examine the mug, they are informed of the possibility of trade. We think of this opportunity as a "surprise." Surprises are formalized by assuming a binary $Q = (D, 1-\epsilon; D', \epsilon)$, where ϵ is small, D is the choice set the person expects, and D' is the decision she is unexpectedly thrust into. Thus, we think of a surprise opportunity as an ex ante low-probability event that happens to materialize. As a consequence, a person's expectations are mostly based on a different choice set than the one she actually chooses from. In endowment-effect experiments, where the opportunity for trade is typically not highlighted until after subjects are given ownership, preferences will be determined by ownership status. In less extreme form, a similar logic applies to experiments where trades are carried out only probabilitically, because in these situations subjects are likely to retain their status quo even if they are aware of the trading opportunities and would like to take advantage of them.

4 Equilibrium Shopping

In this section we explore some of the implications of our model for the market behavior of consumers, where our theory has important implications for how a consumer's valuation for a good is endogenously determined by market conditions and her own anticipated behavior. While we believe the issues in this paper have far less *direct* influence on firm behavior (firms are less likely to be subject to gain-loss effects), market response to the consumer phenomena we discuss could have a profound indirect effect on the behavior of firms, marketers, and salespeople.[23]

Suppose there are two dimensions of choice: shoes and money, with $m(c) = c_1 + c_2$. To isolate the consequences of loss aversion, we assume that μ satisfies A3$'$: it is two-piece linear with $\mu'_+(0) = \eta$ and $\mu'_-(0) = \eta\lambda > \eta$. In this formulation, η can be interpreted as the weight a consumer attaches to gain-loss utility, and λ is her "coefficient of loss aversion". Normalizing the consumer's current endowment to $(0, 0)$, we first consider whether she will buy a single pair of shoes at price p.

[23] See Heidhues and Kőszegi (2004) for an analysis of a profit-maximizing monopolist's response to consumer loss aversion.

Suppose the consumer expects to buy the shoes. Then her utility from buying is $1-p$, and her utility from not buying is $\eta(p-\lambda)$. Thus, she will buy the shoes if $p \leq \frac{1+\eta\lambda}{1+\eta} \equiv p_{max}$. By contrast, if she expects not to buy the shoes, her utility from buying is $1-p-\eta\lambda p+\eta$, and her utility from not buying is zero. Thus, she does not buy if $p \geq \frac{1+\eta}{1+\eta\lambda} \equiv p_{min}$. More generally, if the consumer expects to buy with probability $q \in [0,1]$, the utility from buying is $1-p+(1-q)\eta(1-\lambda p)$, and her utility if she does not buy is $q\eta(p-\lambda)$. Thus, she is indifferent between buying and not buying if $q = \frac{(1+\eta\lambda)p-(1+\eta)}{\eta(\lambda-1)(p+1)}$.

These observations in turn mean that there are three pertinent ranges of the market price. For $p > p_{max}$, the unique equilibrium is not to buy the shoes. For $p < p_{min}$, the unique equilibrium is to buy it. But for prices $p \in (p_{min}, p_{max})$, there are two pure-strategy personal equilibria—one where she for sure buys, one where for sure she doesn't.[24] In this range, therefore, the consumer's expectations are *self-fulfilling*: she buys the shoes if and only if she expects to. Thus, in contrast to standard utility theory, her demand is not uniquely determined by her preferences. Intuitively, if the consumer expects to get the shoes, she feels a loss if she does not buy them, and the money she saves in the process is coded as a gain. Being more sensitive to losses, she is drawn to buying. On the other hand, if she expects to keep the money, buying results in a loss of money and a gain in shoes. Thus, she is inclined toward keeping the money.

The consumer's consumption utility combines with her universal gain-loss function to determine what the range of multiple equilibria is. Tversky and Kahneman (1991) suggest that in most domains where sizes of losses and gains can be measured, people value moderate losses roughly twice as much as equal-sized gains. Given Proposition 2, $\lim_{x \searrow 0} \frac{v_k'(-x)}{v_k'(x)} = 2$ is equivalent to $\frac{1+\eta\lambda}{1+\eta} = 2$, which implies that the range of multiple equilibrium prices is $[\frac{1}{2}, 2]$. That is, the indeterminacy ranges from a price equal to half of the consumer's intrinsic valuation for the shoes to twice of it.[25]

[24] In addition, for $p \in (p_{min}, p_{max})$, there is a unique mixed-strategy equilibrium as well. This equilibrium is unstable in the sense that adaptive dynamics will lead the agent away from this equilibrium (and toward one of the pure ones). This unstable mixed-strategy equilibrium leads to counterintuitive comparative statics: as p increases, the mixed strategy requires that the agent buy the good with higher probability.

[25] The multiple-equilibrium phenomenon, and the fact that it can occur over a wide range of prices, has important implications for marketing and sales techniques. In addition to classical objectives, it becomes crucial for marketers to make sure their clients play the "right" equilibrium, willing to buy the product at relatively high prices.

Reference-dependent preferences can lead to even more interesting and surprising behavior when the consumer is uncertain about the price she will pay. As the above example demonstrates, for a price of p_{max}, it is an equilibrium for the consumer to buy the shoes. Now suppose that she faces a possibly lower, uncertain price: with probability one-half, the price is $p_H = p_{max}$, and with probability one-half, it is $p_L > 1$, where $p_L < p_H$.

Though the price has decreased (in a stochastic-dominance sense), it is now not an equilibrium for the consumer to buy the shoes with probability one. If that was her reference point, then her utility from buying in the high-price state would be

$$1 - p_H - \frac{1}{2}\eta\lambda(p_H - p_L).$$

The second part of this expression is the loss associated with comparing the high price to the lower one the consumer could have gotten. The consumer's utility from not buying is

$$\frac{1}{2}\eta(p_H + p_L) - \eta\lambda.$$

Relative to buying, not doing so leads to (average) savings of $\frac{1}{2}(p_H + p_L)$, and to a loss of the good. By the definition of $p_H = p_{max}$, these expressions imply that the consumer prefers not to buy when the price is high. Is it an equilibrium for her to buy if and only if the price is low? Suppose her expectation is to do so. Then her utility from buying when $p = p_L$ is

$$1 - p_L + \frac{1}{2}\eta(-\lambda p_L + 1), \tag{6}$$

and her utility from not buying is

$$\frac{1}{2}\eta(-\lambda + p_L). \tag{7}$$

For $p_L > 1$, she prefers not to buy the shoes.

While at the deterministically high price it is an equilibrium for the consumer to buy the shoes, at these stochastically lower prices it is the *unique* equilibrium never to buy! The intuition for this violation of the law of demand is the following. Once there is a possibility of acquiring the shoes at a cheaper price, the consumer would experience a loss from paying the higher price instead. Due to this "comparison effect," she does not buy at the high price. But as our multiple-equilibrium example demonstrates, a decrease in the probability with which the consumer expects to buy the shoes decreases her willingness to pay for it. Therefore, if the lower price is still sufficiently high, she never buys them.[26]

The comparison effect is crucial for a price drop to decrease demand. That is, the lower price distribution has to be stochastic for our example to work: if it is an equilibrium to buy at a given (stochastic or deterministic) price, it is also an equilibrium to buy at any lower deterministic price.

In standard consumer theory, a consumer's reservation price for an item does not depend on the ex ante distribution of prices. As the above example illustrates, with reference-dependent preferences the distribution can radically and systematically change the reservation price.[27]

In general, consumer behavior is considerably more complex in our framework than in the standard one. We now turn to a more comprehensive analysis of the consumer's behavior as a function of the distribution of prices she is facing, and derive a few properties of personal equilibrium.

Suppose F is the cumulative distribution function of possible prices for the shoes, with support $[\underline{p}, \overline{p}]$. For any given reference point, it is clear that there is a price p_r (not necessarily in $[\underline{p}, \overline{p}]$) below which the consumer buys the good, and above which she does not. Consequently, equilibrium behavior also has this property. To solve for equilibrium, then, consider, for any p_r and $q \in [0, 1]$, the distribution of outcomes $G_b(p_r, q)$ induced by buying the good with probability one if the price is less than p_r, and with probability q if the price is exactly p_r (and not buying if its price is greater than p_r). We define the correspondence $B(p_r)$ in the following way. For any $G_b(p_r, q)$, consider the net utility associated with buying the good at p_r, when the reference point is $G_b(p_r, q)$:

$$U\left((1, -p_r)|G_b(p_r, q)\right) - U\left((0, 0)|G_b(p_r, q)\right). \tag{8}$$

[26]The presence of multiple equilibria is not necessary for a violation of the law of demand to occur. Due to the comparison effect, it could be the case that for a distribution of prices the unique equilibrium is to buy with probability one, but for a stochastically lower distribution the unique equilibrium is to buy with probability less than one. To construct such an example, both distributions need to be stochastic.

[27]A particular implication of this property of our model is that value revelation mechanisms that are traditionally considered "incentive compatible"—such as a second-price auction with private values or the Becker-DeGroot-Marschak (1964) procedure—do not in fact reveal any intrinsic valuation, but are sensitive to participants' expectations about the outcome.

$B(p_r) \subset \mathbb{R}$ is the set of these net utilities, as we vary q. Note that $B(p_r)$ is single-valued unless p_r is an atom of F. Clearly, p_r^* is an equilibrium reservation price if $0 \in B(p_r^*)$. It is clear that $B(p_r)$ is upper semi-continuous and convex-valued. Furthermore, $B(0) > 0$, and $\lim_{p_r \to \infty} B(p_r) = -\infty$. Thus, there is at least one equilibrium.

For an atomless distribution, it is easy to derive the function $B(p_r)$. If the consumer expects to buy the good below prices p_r, her utility from buying at p_r is

$$\underbrace{1 - p_r}_{\substack{\text{cons.}\\ \text{utility}}} + \underbrace{(-\eta\lambda F(p_r)\,(p_r - E[p|p < p_r]))}_{\substack{\text{gain-loss utility}\\ \text{relative to buy}\\ \text{states}}} + \underbrace{\eta(1 - F(p_r))(1 - \lambda p_r)}_{\substack{\text{gain-loss utility}\\ \text{relative to not buy}\\ \text{states}}}.$$

The first term is the consumption utility derived from buying the good. The second term is the consumer's gain-loss utility from comparing a purchase at price p_r to other purchases, which are all at lower prices. The third derives from comparing a purchase to not purchasing the good. This results in a loss of p_r, and a gain of the good, which is valued at 1. Not buying the good offered at p_r results in a total utility of

$$\eta F(p_r)(E[p|p < p_r] - \lambda),$$

which derives from comparing not buying the good to the possibility of buying. This leads to a sense of gain from saving the average purchase price of the shoes, and a sense of loss due to not having the shoes. Combining these two expressions, we get

$$B(p_r) = 1 - p_r - \eta(\lambda - 1)F(p_r)\,(p_r - E[p|p < p_r]) + \eta(1 - F(p_r))(1 - \lambda p_r) + \eta F(p_r)(\lambda - p_r). \quad (9)$$

To illustrate how small changes in the distribution can affect demand, we assume specifically that p is uniformly distributed on $[\underline{p}, \overline{p}]$. Since the distribution has no atoms, $B(p_r)$ is a function. Equation 9 implies

$$B(p_r) = \begin{cases} 1 - p_r + \eta - \eta\lambda p_r & \text{for } p_r \le \underline{p} \\ 1 + \eta\lambda - (1+\eta)p_r - \eta(\lambda - 1)\left(p_r - \frac{\bar{p}+\underline{p}}{2}\right) & \text{for } p_r > \bar{p} \\ 1 - p_r - \eta(\lambda - 1)\frac{p_r - \underline{p}}{\overline{p} - \underline{p}} \cdot \frac{p_r - \underline{p}}{2} + \eta\frac{\overline{p} - p_r}{\overline{p} - \underline{p}} \cdot (1 - \lambda p_r) + \eta\frac{p_r - \underline{p}}{\overline{p} - \underline{p}} \cdot (\lambda - p_r) & \text{for } p_r \in [\underline{p}, \bar{p}]. \end{cases} \quad (10)$$

Equation 10 is a quadratic function of p_r. Simple arithmetic shows that the coefficient on p_r^2 is $\frac{\eta(\lambda-1)}{2(\bar{p}-\underline{p})} > 0$, and the coefficient on p_r is $\frac{\eta(\lambda-1)}{\bar{p}-\underline{p}} - 1 - \eta\lambda$.

We investigate the behavior of equilibria when $\underline{p}$ is around p_{min}, under the assumption that the quadratic part of $B(p_r)$ is increasing. This is equivalent to the following condition:

$$B'(p_{min}) = \frac{\eta(\lambda-1)(p_{min}+1)}{\bar{p}-\underline{p}} - 1 - \eta\lambda > 0,$$

which is true if $\bar{p} - \underline{p} < 1 - \frac{(1+\eta)^2}{(1+\eta\lambda)^2}$. In that case, as long as $\underline{p} \geq p_{min}$, $B(p_r)$ has a zero at p_{min} (since the first part of the curve has a zero at p_{min}). That is, it is an equilibrium for the consumer never to buy the good. Now suppose $\underline{p} < p_{min}$, but $\underline{p} \approx p_{min}$. Then, no matter how little $\underline{p}$ slips below p_{min}, the unique equilibrium is to buy the good with probability 1.

The intuition behind this discontinuity is the following. Once there is a chance for the price to drop below p_{min}, in any equilibrium the consumer will want to buy the good with at least some probability. But expecting some probability of buying, she now feels a small loss if she does not get the good. This "attachment effect" increases the reservation price to at least slightly above p_{min}. This increases the probability with which she expects to buy, increasing the attachment and thus worsening the pain of loss in case she does not, again increasing the probability of purchase, and so on. If $\bar{p} - \underline{p}$ is sufficiently small, this self-reinforcing mechanism is sufficiently strong for the equilibrium to unravel. Although the comparison effect is operational in this example, it is outweighed by the attachment effect. Once again, this is assured by the fact that $\bar{p}-\underline{p}$ is sufficiently small (so that the loss the consumer suffers when paying the high price is not too great).

We can use experimental evidence to calibrate the strength of the attachment effect as well. If we take $\frac{1+\eta\lambda}{1+\lambda} = 2$ as before, then $\bar{p} - \underline{p}$ could be close to $\frac{3}{4}$ (of the consumer's valuation for the good), and unraveling would still occur. This implies that with $\underline{p}$ slightly above one-half (p_{min}), the consumer does not necessarily buy the good at that price. But when $\underline{p}$ slips below one-half, in the unique resulting equilibrium she buys it even at a 150 percent higher price.

In both of the unraveling examples above, uncertainty leads to a unique equilibrium. The tendency for sufficient uncertainty to lead to a unique equilibrium is a general phenomenon:

Proposition 3 *For all η and λ, there exists $\varepsilon > 0$ such that for all non-atomic price distributions*

F for which $f(\cdot) < \varepsilon$ everywhere, the personal equilibrium is unique.

This result is potentially very important for applications of our model where the indeterminacy of consumer behavior due to the multiple-equilibrium phenomenon makes analysis difficult, and where no obvious equilibrium selection rule presents itself.

While uncertainty leads to a unique equilibrium in both of our examples, the examples are very different in at least one way: in the first, a probabilistic price decrease leads to a decrease in demand, while in the other it leads to an increase. The following proposition summarizes the principles behind the two examples.

Proposition 4 *Suppose F and F' are nonatomic distributions, and p_r^* is a personal-equilibrium reservation price when the price distribution is F. Then,*

1. *If $F'(p_r^*) \leq F(p_r^*)$, and $\int_{\underline{p}}^{p_r^*} p dF'(p) < \int_{\underline{p}}^{p_r^*} p dF(p)$, then there is a $p_r^{**} < p_r^*$ that is a personal-equilibrium reservation price when the price distribution is F'.*

2. *If $F'(p_r^*) > F(p_r^*)$, and $\int_{\underline{p}}^{p_r^*} p dF'(p) \geq \int_{\underline{p}}^{p_r^*} p dF(p)$, then there is a $p_r^{**} > p_r^*$ that is a personal-equilibrium reservation price when the price distribution is F'.*

Proposition 4 can be thought of as summarizing the effect of possible sales (probabilistic price decreases) on consumers' reservation prices, and consequently demand. We can distinguish two situations. If the price drops down from a level at which the consumer would have bought anyway, then the existence of the sale has the paradoxical effect of decreasing demand (Part 1 of the proposition). This is due to the comparison effect. The reservation price is by definition the highest price the consumer is ever going to pay for the good, so buying at that price necessarily feels like a loss when compared to other possible purchase prices. And the lower the other possible prices, the greater the consumer's sense of loss from buying, pushing down the reservation price.

When the price drops from above the decisionmaker's reservation price to below it, however, the price decrease can increase demand. This in itself is not too surprising, since the probability that the price is below the reservation level increases. But by Part 2 of Proposition 4, such a price drop can have the additional effect of increasing the reservation price, and thereby further increasing demand. This "multiplier" is the attachment effect at work. Since the consumer now expects to buy with a higher probability, she is more attached to the idea of having the good, increasing her reservation price.[28]

In interpreting these results, some care must be taken in identifying which price changes should indeed be modeled as random prices. A stochastic price in our model corresponds to true price uncertainty, where the consumer forms expectations before finding out the price. This is different from price variations that consumers can predict and incorporate into their expectations. On Memorial Day sales, for example, prices at many stores are significantly lower than at most other times of the year. Nevertheless, since consumers know about and expect these sales, the appropriate way to specify their decision problem is to assume a price distribution weighted heavily toward low prices. Therefore, our theory predicts that consumers would in general respond differently to these sales than to less predictable price drops.

The logic behind Proposition 4 could also shed light on non-equilibrium situations in which buyers believe that the price is lower than it actually is. For example, car dealers sometimes use the strategy of "throwing a lowball," in which they promise a really low price, and then attempt to raise it back up once the consumer gets used to the expectation of buying.[29] According to our model, this strategy may backfire, because (due to the comparison effect) the consumer may refuse to buy at the high price if she was expecting the low one. The strategy is successful, however, if it engages the attachment effect in a consumer who would not buy at the high price.

The above analysis applies to the agent's reactions to price changes in personal equilibrium. In this sense, it is about "long-run" responses—the consumer's eventual demand once her expectations adjust. We can distinguish this from "short-run" responses to surprise price changes—her demand before expectations adjust. The comparison of short-run and long-run price responses leads to

[28]Note that in order for the attachment effect to create a multiplier, the price cannot drop too much below the reservation price. If that were to happen, the comparison effect would dominate the attachment effect, and the price decrease might once again lead to a decrease in demand.

[29]See Cialdini (1993) for a social-psychological discussion of such sales tactics.

further stark implications of our model in consumer behavior. To avoid the indeterminacy of multiple equilibria, we we analyze this issue assuming that the price is sufficiently risky for the equilibrium to be unique. We consider a downward shift in the distribution of prices, and ask whether the short-run or long-run demand response is greater. Formally, a downward shift from a given price distribution F is modeled as follows. For any positive real z, consider the distribution F^z defined by $F^z(p) = F(p+z)$. Denote the equilibrium reservation prices for price distributions F and F^z by p_r^* and p_r^{z*}, respectively. Similarly, define the function $B^z(p_r)$ according to Equation 9, with the distribution of prices replaced by F^z.

Since the consumer's reservation price remains fixed for any short-run price change, the question of whether the long-run price elasticity is larger than the short-run one is equivalent to asking whether $p_r^{z*} > p_r^*$. If $p_r^{z*} > p_r^*$, the consumer will be more willing to buy the good after "getting used to" the new price distribution than beforehand. But if the inequality goes the other way, her long-run price elasticity is lower than than her short-run one.

Simple calculus implies that

$$\left.\frac{\partial B^z(p_r^*)}{\partial z}\right|_{z=0} = \eta(\lambda-1)\left(f(p_r^*)(1+p_r^*) - F(p_r^*)\right).$$

If this expression is greater than zero, then $p_r^{z*} > p_r^*$ for small positive price shifts z; otherwise, the opposite is the case. Intuitively, there are two opposing effects that contribute to the difference between short-run and long-run elasticities. First, as the consumer becomes to expect lower prices, she attaches a higher probability to purchasing the good. Therefore, she feels more of a loss if she does not buy it, and less of a loss in money if she does. This is analogous to the attachment effect, and increases her long-run demand relative to the short-run one. On the other hand, as the consumer gets used to the low prices, prices that seemed to be reasonable beforehand now invoke more of a sensation of loss. This comparison effect decreases her long-run demand.

The consumer's probability of buying the good ($F(p_r^*)$) determines the importance of the comparison effect, since this effect derives from the comparison of the realized price to other possible purchase prices. And the consumer's price responsiveness (which is proportional to $f(p_r^*)$) determines the importance of the attachment effect, since this effect derives from a change in the consumer's anticipated probability of purchase. Thus, it is clear that either effect could dominate.

In particular, the comparison effect dominates if $F(p_r^*)$ is large relative to $f(p_r^*)$—if the consumer's short-run price elasticity is small.

These predictions distinguish our model from intertemporal models with habit-forming or experience goods. In habit-formation models, typically the long-run elasticity of consumption is higher than the short-run elasticity, because in the long run people get used to their new level of consumption. Our model implies not only that the opposite is possible as well, it provides a simple testable condition for when it is likely to be so.

We conclude our market analysis by characterizing the behavior of *sellers* facing a distribution of prices with probability measure F, and compare the behavior of buyers and sellers. Unlike in the analysis so far, we are primarily motivated by the endowment effect rather than real markets per se; marketplace sellers are typically not individuals with endowments of goods. We assume that the seller in question has the same utility function (including the same valuation for the shoes) as the buyer. In a standard setting, this would imply that the seller and the buyer have the same reservation price.

As in the case of buyers, in equilibrium the seller follows a reservation price policy, selling the good when the price is above some given p_r^{**}. For simplicity, we assume that the distribution of prices is non-atomic, and look for the equilibrium reservation price in a similar way as for buyers. When the seller expects to sell at prices greater than p_r, her utility from selling at price p_r is

$$p_r - 1 - \eta\lambda\left(1 - F(p_r)\right)\left(E[p|p > p_r] - p_r\right) + \eta F(p_r)(p_r - \lambda).$$

The first term in the expression is simply the consumption utility from the transaction. The second term comes from comparing the consumer's consumption to states in which she sells. Since in those states she sells the good at higher prices, she experiences a loss. And the last term derives from comparing her consumption to states in which she does not sell. This leads her to assess a gain in money equal to the price, and a loss of the good. The seller's utility from holding on to the shoes in the same situation is

$$-\eta\lambda\left(1 - F(p_r)\right)E[p|p > p_r] + \eta\left(1 - F(p_r)\right).$$

Thus, the net utility from holding to the good relative to selling it is

$$(1+\eta)(1-p_r)+\eta(\lambda-1)F(p_r)-\eta(\lambda-1)(1-F(p_r))p_r$$

As an example, consider the case we have analyzed above for buyers: $p \sim U[\underline{p}, \bar{p}]$. Assume further that $\frac{\underline{p}+\bar{p}}{2}=1$. Notice that $S(1)=0$ in that case, so it is an equilibrium for the seller to adopt the value of the good as her reservation price. It is easy to show that $B(1)<0$, so there is an equilibrium in which the buyer has a reservation price less than 1. This result also generalizes:

Proposition 5 *For any personal equilibrium reservation price p_r^* of the buyer, there is a personal equilibrium reservation price $p_r^{**} \geq p_r^*$ of the seller, and if $F(p_r^*)>0$, then $p_r^{**}>p_r^*$. Conversely, for any personal equilibrium reservation price p_r^{**} of the seller, there is a personal equilibrium reservation price $p_r^* \leq p_r^{**}$ of the buyer, and if $F(p_r^{**})>0$, then $p_r^*<p_r^{**}$. In particular, if reservation prices are unique, the buyer's is lower.*

Our model therefore endogenously generates the endowment effect from the roles of the players, even though it does not assign any special role to initial endowments. The force that pushes down the buyer's reservation price is the comparison effect from above: if she pays the reservation price for the good, she experiences a loss relative to lower prices, and she can avoid this loss by not buying. When the seller sells at the reservation price, she only makes a gain (as opposed to avoiding a loss) relative to the low-price states, in which she does not sell. Of course, she experiences a loss relative to higher-price states; the reservation price is the lowest price she could have sold at. But crucially, the seller *does not affect* this loss by lowering her reservation price. Whether or not she sells at price p_r, she experiences a loss relative to states in which the price is greater than p_r.[30]

[30] The endowment effect and Proposition 5 seem to indicate a suppression of market trade due to the gap between buying and selling prices. But importantly, our theory predicts that neither the findings of endowment-effect experiments, nor the conclusions of Proposition 5 will generalize to real market settings. In the marketplace, it is likely that gains from trade among those who participate in the market are positive. In that case, sellers will expect to sell, and the buyers to buy, which reverses the gap between the buying and selling price.

5 Hedonic Dimensions

A central part of our model is the set of consumption dimensions on which gain-loss utility is defined. Because our theory (or any theory) of reference dependence would make very bad predictions in some situations if applied literally to physical consumption dimensions, in this section we clarify how best to think of these dimensions. To illustrate the problem, consider a decisionmaker making choices over Tropicana and Florida's Natural premium orange juices, two separate consumption goods she really enjoys but can barely distinguish by taste or nutritional content. Would she be willing to trade six ounces of Tropicana juice she had expected to consume for eight ounces of Florida's Natural juice? If the trade were coded as a loss in the Tropicana dimension and a gain in the Florida's Natural dimension, under usual parametrizations our model predicts she would not.

If the person does not care what kind of orange juice she consumes, however, she would presumably accept the trade. Psychologically, it is likely that the two brands would be treated as a single dimension in gain-loss utility, so that the trade would be assessed as a gain of two ounces of premium orange juice. More generally, most consumption goods are mixtures of different hedonic attributes which could be subject to gain-loss effects. Orange and grapefruit juices may satisfy a person's thirst to the same extent, but she may find their taste quite different. Hence, giving up orange juice that she expected to receive for grapefruit juice would not generate a loss or gain in terms of satisfying thirst, but would generate a loss in terms of orange taste and a gain in terms of grapefruit taste. Hence, a full account of the consumer's choice includes breaking up this two-good choice into a three-dimensional decision, consisting of assessing outcomes in terms of orangeness, grapefruitness, and juiciness. Reconceptualizing the choice thusly would not affect standard consumer theory—all that matters is the net utilities for the two goods—but very much affects the predictions of any theory of reference dependence where the dimensions used to evaluate outcomes are treated separately.

These examples suggest that the appropriate way to think about our model is to assume that decisionmakers evaluate gains and losses over hedonic dimensions. That is, we can take any K-dimensional consumption decision and re-write it as an H-dimensional choice problem by positing that a vector of the K goods yields a vector of consumption levels in the H hedonic dimensions.

This rewriting allows different kinds of orange juice to be treated as a single dimension, as well as a single good to affect more than one hedonic dimension. Once transformed, the model of the previous sections can be applied directly.[31]

The relevant hedonic dimensions can be related to the original dimensions in many ways. It could be, for instance, that a person's preferences over peanut butter, jelly, and bread are determined solely by the number of peanut butter and jelly sandwiches produced. We would then want to translate the levels of the dimensions (p, j, b) into a single dimension $c_1 \equiv Min[p, j, b]$. If the consumer also cares about a "calorie dimension", a second hedonic dimension of $c_2 \equiv p + j + b$ could be added.

In this framework, two goods that are perfect substitutes in the classical sense—meaning that the marginal utility from consuming one does not depend on the consumption level of the other—may or may not be considered substitutes when assessing gains and losses of combinations of these goods. The extent to which they are integrated depends on the extent to which a person considers their consumption to be hedonically similar experiences. To illustrate this, suppose a moviegoer's utility from any given film depends on how it contributes to each of four hedonic dimensions: acting, plot, cinematography, and music. If she cares equally about the four dimensions, and has utility from money, as well, then her consumption utility for a given movie can be given by $m(c) = c_1 + c_2 + c_3 + c_4 + c_5$.

Now suppose that Movies A and B each provide one unit of consumption in two of the relevant hedonic dimensions, and zero units of consumption in the other two. Movies A and B are then perfect substitutes in the classical sense. But "gain-loss substitutability", as reflected in the consumer's willingness to trade off one for the other, when she was expecting to consume exactly one of them, also depends on whether the movies provide utility in the same hedonic dimensions. Suppose Movie A provides positive consumption in the acting and cinematography dimensions, and that the decisionmaker is expecting to watch Movie A at price p_A (so $c_A = (1, 1, 0, 0, -p_A)$), but (unexpectedly to her) she gets an opportunity to switch to Movie B at price p_B. What p_B would

[31]Note that H may be bigger or smaller than K. In the case of different kinds of orange juice, $K = 2$ and $H = 1$; in the case of orange juice vs. grapefruit juice, $K = 2$ and $H = 3$.

induce a switch for the moviegoer? If $c_B = (1, 1, 0, 0, -p_B)$, the consumer will prefer Movie B if it is cheaper, no matter how little the difference in price. If the two movies differ in one dimension, however, with $c_B = (0, 1, 1, 0, -p_B)$, then $p'_B < p_A$ would be needed to induce a switch; if they differ in two dimensions, with $c_B = (0, 0, 1, 1, -p_B)$, a yet lower price $p''_B < p'_B$ would be needed.[32]

The premise that the relevant hedonic dimensions may differ from goods dimensions means that substantive psychological assumptions about these dimensions are needed in applying our model. This conceptual indeterminacy in translating consumption utility into a full specification of utility is one of the biggest caveats to our claim in the introduction that we prescribe a fully-determined way to translate any classical model into a reference-dependent one.

It is worth emphasizing, however, that this last statement is (merely) about the ability to derive our model from a reference-independent one where researchers believe they already know the appropriate consumption utility functions. In any environment where expectations by consumers can be measured, inferred, or induced, standard revealed-preference analysis of the type used ubiquitously in "reference-independent economics" can be used to fully identify utility functions $u(c|r)$.

Indeed, because behavior depends on which hedonic dimensions a person considers as separate, revealed-preference analysis can also be used to directly identify hedonic dimensions. In particular, the above analysis of the consumer's willingness to switch between two consumption subsitutes provides a way to determine—at least in principle—whether she considers two dimensions as hedonically separate. Suppose we start from a "maximal" set H_0 of dimensions of experience that a person might plausibly care separately about. Then, there is a simple procedure to extract, from the consumer's behavior, whether $h_1, h_2 \in H_0$ should be specified as separate dimensions in gain-loss utility. First, notice that we can find consumption vectors c^1 and c^2 that give rise to the same consumption utility, and that only differ in dimensions h_1 and h_2: we promise a person zero consumption in all dimensions, and then offer her various choices that differ only in dimensions h_1 and h_2, until we pin down a pair of vectors she is indifferent between.[33] Then, we promise the

[32]More precisely, p'_B solves $p_A - p'_B + \mu(p_A - p'_B) = -\mu(-1) - \mu(1)$ and p''_B solves $p_A - p''_B + \mu(p_A - p''_B) = 2(-\mu(-1) - \mu(1))$.

[33]With rational expectations, it is of course impossible to literally surprise the agent in this way. But just as

person the consumption vector c^1, and then offer her the opportunity to exchange it for c^2 and a little bit of money. If she refuses the exchange, h_1 and h_2 are truly separate hedonic dimensions. If she accepts, they are not. Repeating this procedure, we obtain a partition of H_0, where each member of the partition is revealed to be a set of hedonically inseparable dimensions. The partition should be used as the basis of the reference-dependent model.[34]

Of course, any re-writing of the dimensions should preserve the substantive consumption utility (as opposed to gain-loss utility) of the original formulation, and to apply our model, we need to maintain the linear separability of the dimensions. It is here, however, that we posit the psychological hypothesis that the dimensions by which people assess gains and losses are indeed separable. That is, if a person's utility for goods x and y is not additively separable, then she will not assess gains and losses separately in terms of those goods. Psychologically, the very reason that x and y are not separable in consumption utility is that there is a hedonic dimension to which both contribute in some way. A left shoe is a complement to a right shoe because both are necessary to the experience of having a pair of shoes to wear, which is what a consumer cares about in the end. And a Chinese meal is a partial substitute to an Italian meal because both can satisfy a person's need to be satiated; that the utility from eating one meal depends inseparably on the total number of meals is an indicator in our interpretation that calorie intake needs to be considered as a separate dimension in gain-loss evaluation.

We believe it is transparent that any plausible general model of reference-dependent preferences must be defined with respect to dimensions that treat hedonically similar experiences as single dimensions. More speculatively, we conjecture that there are also circumstances that are best conceptualized by assuming that even hedonically initially distinct dimensions become similar for the decisionmaker in evaluating gain-loss utility. Consider a consumer who, over a long period of

discussed in Section 3, we can get "arbitrarily close." Here, an agent can be told that she will have zero consumption in all dimensions with probability $1-\epsilon$, and be able to choose with very low probability ϵ.

[34]It bears emphasizing, however, that the identification of the relevant hedonic dimensions are merely means of predicting which consumption goods are complete or partial "gain-loss substitutes" as reflected in the function $u(c|r)$, and of psychologically interpreting this function. Once $u(c|r)$ is empirically identified, our model is fully specified, with no further assumptions about the relevant hedonic dimensions needed for analysis.

time, randomly either plays basketball or goes out on Friday night. Although these two experiences are intitially clearly separate hedonically, we hypothesize that the consumer's experience in trading them off leads to their integration. That is, after a while the consumer gets used to having her losses in one activity being offset by gains in the other, to the extent that she does not experience giving up one for the other as a loss.[35]

6 Discussion and Conclusion

Our goal in this paper was to put forward a fully specified model of reference-dependent preferences that can accommodate existing evidence and, most importantly, be applied to a wide range of economic situations. The centerpiece of our model is the proposal that a person's reference point is her recent probabilistic beliefs about the outcomes she is going to get. Thus, for example, if she expects improvements in her circumstances, and these changes fail to occur, she experiences a painful sensation of loss, even if she has retained or improved on her status quo. Indeed, our model provides an avenue to study an intuition about the strong role that expectations play in employee satisfaction with wages; it predicts both a status quo bias in stagnant environments, and a taste for improvement in environments where workers have become accustomed to improvement.[36]

By assuming rational expectations, our theory fully endogenizes the determination of reference points. In fact (with the important caveat that applying our theory requires a modeler to correctly specify the dimensions over which gain-loss utility is defined) we have provided a way to translate a classical model into a reference-dependent one, and demonstrated such applicability in consumer behavior. Since the theory emphasizes the malleability and endogenously determined nature of

[35]In general, the integration of different hedonic dimensions due to experience with trading them off may be an important phenomenon in any consumer society: as savvy consumers, we are used to getting the best possible deal each time, and thus to trade off goods we might have expected buying but are not available cheaply with those we might not have expected to get but are selling at a bargain.

[36]On a larger scale, this idea is consistent with Davies's (1962) influential theory of political revolutions. Davies claims that revolutions invariably occur when people have come to expect improvements in conditions, and an economic downturn occurs. An "intolerable gap" then develops between the changes that the people expect and what they actually get.

valuations for consumption goods, it is a fruitful arena in which to study competition between firms and psychological aspects of sales techniques (such as the low-balling of car buyers mentioned above), marketing, and advertising.[37]

There are some important economic situations to which the current version of our model does not apply, but to which we believe alternative models using similar principles would apply. As we have stressed, the psychological hypothesis behind our formulation is that a person's reference point, and thus her preferences over consumption outcomes, depend on lagged expectations. As a consequence, when the person makes her consumption decision, she treats the reference point as fixed, giving rise to our personal-equilibrium concept. In some important economic situations, however, a considerable amount of time passes between the person's decision and the time of consumption. In these situations, the reference point to which consumption is eventually compared will presumably be based on the decision made; insofar as the decisionmaker anticipates this, she would *not* treat the reference point as fixed. A good example for this kind of decision problem is the purchase of extended warranties and other types of insurance: whether or not a good breaks is typically realized long after the decision. Incorporating such possibilities into a rational-expectations model requires a modification of our theory that allows a person's reference point to be determined after her decision. In its current form, therefore, our model is applicable mostly to situations where the decision leads relatively quickly to consumption.

Finally, because our framework assumes fully rational utility maximization, it does not incorporate two aspects of loss aversion that evidence strongly suggests are important. First, we have not seriously confronted issues of how a person conceptualizes a given decision problem, and, in particular, which future, contemporaneous, or potential decisions she takes into consideration when thinking about the current one. While all economic models are sensitive to assumptions regarding such "choice bracketing," the issue is especially acute in models such as ours. What other decisions

[37]We also suspect that our model may help intepret, improve, and bring into economic analysis the vast marketing literature on two topics: Brand loyalty (see, e.g., Aaker 1991,1995) and reference-price effects (see., e.g., Bidwell, Wang, and Zona (1995) and Putler (1992)). While—especially in understanding brand loyalty—other factors seem important, these two phenomena are plausibly related to the role of reference points in both purchases of goods and the prices paid.

a person considers as integral to the current one will affect our prediction about what she considers to be gains and losses.[38] Substantial evidence indicates that people *narrowly bracket*—they do not integrate decisions at hand with other ones, even when it is clear that they would increase their utility by doing so. One cannot fully understand reference-dependent preferences without understanding these bracketing phenomena.[39] Second, there is also evidence that the observed degree of loss aversion does reflect true reference-level effects on well-being, but involves misprediction about those effects. One does feel a little loss at losing a mug. That, by all intuition and evidence, is a real hedonic experience, and making choices reflecting that real hedonic experience is partly rational. But as interpreted by Kahneman (2001) and Loewenstein, O'Donoghue, and Rabin (2003), people seem to over-attend to this experience because they ignore that the sensation of loss will pass very quickly—behaving as if they would spend much time longing for the mug they once had.[40] On both of these accounts, the nature and scope of reference-dependent choices seems to reflect mistakes. Therefore, our model—which assumes rational utility maximation—is incomplete in describing the implications of reference dependence, especially when welfare concerns are involved.

Appendix: Proofs

Proof of Proposition 1.

1. Obvious.
2. Suppose not; that is, suppose $u(c, c') \geq u(c', c')$ and $u(c, c) \leq u(c', c)$. Adding these and

[38]For example, if she completely isolates a purchase decision from others, not buying after expecting to do so necessarily feels like a loss. But if she brackets broadly, she would consider the possibility of "making up" for the loss with other purchases.

[39]For discussions of choice bracketing in various contexts, see, e.g., Kahneman and Lovallo (1993), Benartzi and Thaler (1995), and Read, Loewenstein, and Rabin (1999).

[40]This form of misprediction leads to one way we believe our model over-states the connection beetween between consumption and gain-loss utility. Items that are subject to misprediction about attachment are likely to exhibit a higher gain-loss to consumption utility ratio than those items (e.g., paper clips, toilet paper, or a household furnace) whose purpose is sufficiently utilitarian that people neither become strongly attached to them, nor mispredict their future utility for them.

using the definition of u implies that

$$m(c) + m(c') + n(c|c') + n(c'|c) \geq m(c) + m(c') + n(c|c) + n(c'|c'). \tag{11}$$

Eliminating $m(c) + m(c')$ from both sides and using the definition of n, this reduces to

$$\sum_{k=1}^{K} \left[\mu(m_k(c_k) - m_k(c'_k)) + \mu(m_k(c_k) - m_k(c'_k))\right] \geq 0. \tag{12}$$

By A2, and using that $c \neq c'$, this is a contradiction.

3. We prove that for any $F, F' \in \Delta(\mathbb{R}^K)$,

$$U(F|F) + U(F'|F') \geq U(F|F') + U(F'|F).$$

This is obviously sufficient to establish the claim by contradiction. Furthermore, we prove that the above is true dimension by dimension. Let the marginals of F and F' on dimension k be F_k and F'_k, respectively. Noticing that expected consumption utilities are the same on the two sides, we want to prove that

$$\begin{aligned} & \int\int \mu(m_k(c_k) - m_k(r_k))dF_k(c_k)dF_k(r_k) + \int\int \mu(m_k(c_k) - m_k(r_k))dF'_k(c_k)dF'_k(r_k) \\ \geq \ & \int\int \mu(m_k(c_k) - m_k(r_k))dF_k(c_k)dF'_k(r_k) + \int\int \mu(m_k(c_k) - m_k(r_k))dF'_k(c_k)dF_k(r_k). \end{aligned}$$

Since μ satisfies A3$'$, for any $x \geq 0$ we have $\mu(x) + \mu(-x) = -\alpha x$ for some $\alpha > 0$. Using this and dividing by $-\alpha$, the above becomes

$$\begin{aligned} \frac{1}{2}\int\int |m_k(c_k) - m_k(r_k)|dF_k(c_k)dF_k(r_k) + \frac{1}{2}\int\int |m_k(c_k) - m_k(r_k)|dF'_k(c_k)dF'_k(r_k) \\ \leq \int\int |m_k(c_k) - m_k(r_k)|dF_k(c_k)dF'_k(r_k). \end{aligned} \tag{13}$$

We can give a geometric interpretation to this inequality. The inequality says that the expected distance (measured according to the utility function m_k) between two points chosen randomly according to F_k plus the expected distance between two points chosen randomly according to F'_k is less than twice the expected distance between a point chosen randomly according to F_k and a point chosen randomly according to F'_k.

Consider any point x on the real line. Let $F_k(x) = p$ and $F'_k(x) = p'$. The probability that $m_k(x)$ is on a line segment of two points $m_k(c_k)$ and $m_k(r_k)$, where c_k and r_k are chosen independently according to F_k is $2p(1-p)$. Similarly, the probability that it is between two such points when c_k and r_k are chosen according to F'_k is $p'(1-p')$. And the probability that it is between two such points when c_k and r_k are chosen according to F_k and F'_k, respectively, is $p(1-p')+p'(1-p)$. It is sufficient to prove that

$$p(1-p)+p'(1-p') \leq p(1-p')+p'(1-p),$$

which is true since $(p-p')^2 \geq 0$. Furthermore, if there is any x such that $F_k(x) \neq F'_k(x)$, then the inequality 13 is strict. □

Proof of Proposition 2. Let $v_k(x) = m_k(x) - m_k(0) + \mu(m_k(x) - m_k(0))$. Since m_k is linear for each k,

$$\begin{aligned} u(c|r) - u(r|r) &= \sum_{k=1}^{K} [m_k(c_k) + \mu(m_k(c_k) - m_k(r_k))] - \sum_{k=1}^{K} m_k(r_k) \\ &= \sum_{k=1}^{K} [m_k(c_k - r_k) - m(0) + \mu(m_k(c_k - r_k) - m_k(0))] \\ &= \sum_{k=1}^{K} v_k(c_k - r_k). \end{aligned}$$

A0 and A1 are obviously satisfied. Notice that for any $y > x > 0$,

$$\begin{aligned} v_k(-y) + v_k(y) &= \mu(m_k(y) - m_k(0)) + \mu(-(m_k(y) - m_k(0))) \\ &< \mu(m_k(x) - m_k(0)) + \mu(-(m_k(x) - m_k(0))) = v_k(-x) + v_k(x) \end{aligned}$$

since m_k is increasing and μ satisfies A2. Thus, v_k satisfies A2. A3 is obvious, given the linearity of m_k. Next,

$$\lim_{x \searrow 0} \frac{v'_k(-x)}{v'_k(x)} = \frac{m'_k(0) + \mu'_-(0) m'_k(0)}{m'_k(0) + \mu'_+(0) m'_k(0)} = \frac{1 + \mu'_-(0)}{1 + \mu'_+(0)} < \lambda.$$

This completes the proof. □

Proof of Proposition 3. We can put $B(p_r)$ in the following form:

$$B(p_r) = 1 - p_r - \eta\lambda \int_0^{p_r} (p_r - p) f(p) dp + \eta \int_{p_r}^{\infty} (1 - \lambda p_r) f(p) dp - \eta \int_0^{p_r} (p - \lambda) f(p) dp. \quad (14)$$

We prove that if $f(p_r) \leq \frac{1+\eta\lambda}{\eta(\lambda-1)(p_{max}+1)}$, then $B(p_r)$ has exactly one zero. First, expression 14 is positive for $p_r = 0$, and negative for a sufficiently large p_r, so $B(p_r)$ has at least one zero.

To prove that $B(p_r)$ has exactly one zero, notice that it is negative for any $p_r > p_{max}$. Now $B'(p_r) = -1 - \eta\lambda + \eta(\lambda - 1)(p_r + 1) f(p_r)$, which, by the condition on $f(p_r)$, is negative for any $p_r < p_{max}$. □

Proof of Proposition 4. Let $B_F(p_r)$ be defined by Equation 8 when the distribution of prices is F, and let $B_{F'}(p_r)$ be defined similarly for F'. Under conditions of Part 1 of the Theorem, $B_{F'}(p_r^*) < 0$. Since $B_{F'}(0) > 0$ and $B_{F'}(p_r)$ is continuous, there is a $p_r^{**} < p_r^*$ such that $B_{F'}(p_r^{**}) = 0$. Similarly, under the conditions of part of the Theorem, $B_{F'}(p_r^*) > 0$. Since $\lim_{p_r \to \infty} B_{F'}(p_r) < 0$ and $B_{F'}(p_r)$ is continuous, there is a $p_r^{**} > p_r^*$ satisfying $B_{F'}(p_r^{**}) = 0$. □

Proof of Proposition 5. From the analysis in the text,

$$S(p_r) - B(p_r) = \eta F(p_r)(\lambda - 1)\left(p_r - E[p|p < p_r]\right) \geq 0,$$

and the inequality is strict if $F(p_r) > 0$. Combined with $S(0), B(0) > 0$, and $\lim_{p_r \to \infty} S(p_r), \lim_{p_r \to \infty} B(p_r) = -\infty$, this implies the statement of the theorem. □

References

Aaker, David A. "Building Strong Brands." Simon & Schuster, 1995.

Aaker, David A. "Managing Brand Equity: Capitalizing on the Value of a Brand Name." Simon & Schuster, 1991.

Ariely, Dan; Loewenstein, George and Prelec, Drazen. " 'Coherent Arbitrariness': Stable Demand Curves without Stable Preferences." Quarterly Journal of Economics, February 2003, *118*(1), pp. 73-105.

Barberis, Nicholas; Huang, Ming and Santos, Tano. "Prospect Theory and Asset Prices." *Quarterly Journal of Economics*, February 2001, *116*(1), pp. 1-53.

Bateman, Ian; Kahneman, Daniel; Munro, Alistair; Starmer, Chris and Sugden, Robert. "Is There Loss Aversion in Buying? An Adversarial Collaboration." Working Paper, 2002.

Bateman, Ian; Munro, Alistair; Rhodes, Bruce; Starmer, Chris and Sugden, Robert. "A Test of the Theory of Reference-Dependent Preferences." *Quarterly Journal of Economics*, May 1997, *112*(2), pp. 479-505.

Becker, Gordon M.; DeGroot, Morris H. and Marschak, Jacob. "Measuring Utility by a Single-Response Sequential Method." *Behavioral Science*, July 1964, *9*, pp. 226-232.

Benartzi, Shlomo and Thaler, Richard H. "Myopic Loss Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics*, February 1995; *110*(1), pp. 73-92.

Bidwell, Miles O., Jr.; Wang, Bruce X. and Zona, J. Douglas. "An Analysis of Asymmetric Demand Response to Price Changes: The Case of Local Telephone Calls." *Journal of Regulatory Economics*, November 1995, *8*(3), pp. 285-98

Bowman, David; Minehart, Deborah and Rabin, Matthew. "Loss Aversion in a Consumption-Savings Model." *Journal of Economic Behavior and Organization*, February 1999, *38*(2), pp. 155-78.

Breiter, Hans C.; Aharon, Itzhak; Kahneman, Daniel; Dale, Anders and Shizgal, Peter. "Functional Imaging of Neural Responses to Expectancy and Experience of Monetary Gains and Losses." *Neuron*, May 2001, *30*, pp. 619-639.

Camerer, Colin; Babcock, Linda; Loewenstein, George and Thaler, Richard. "Labor Supply of New York City Cabdrivers: One Day at a Time." *Quarterly Journal of Economics*, August 1997; *112*(2), pp. 407-441.

Cialdini, Robert B. *Influence: The Psychology of Persuasion.* New York: Morrow, 1993.

Clark, Andrew and Oswald, Andrew. "Satisfaction and Comparison Income." *Journal of Public Economics*, September 1996, *61*(3), pp. 359-81.

Davies, James C. (1962) "Toward a Theory of Revolution," American Sociological Review, v. 27 nr. 1, pp. 5-18

Duesenberry, James S. *Income, Saving, and the Theory of Consumer Behavior*. Cambridge, MA: Harvard University Press, 1952.

Frederick, Shane and Loewenstein, George. "Hedonic Adaptation," in Daniel Kahneman, Ed Diener and Norbert Schwarz, eds., *Well-being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation, 1999, pp. 302-29.

Frey, Bruno S. and Stutzer, Alois. "What Can Economists Learn from Happiness Research?" *Journal of Economic Literature*, June 2002, *40*(2), pp. 402-35.

Geanakoplos, John; Pearce, David and Stacchetti, Ennio. "Psychological Games and Sequential Rationality." *Games and Economic Behavior*, 1989, *1*(1), pp. 60-79.

Genesove, David and Mayer, Christopher. "Loss Aversion and Seller Behavior: Evidence from the Housing Market." *Quarterly Journal of Economics*, November 2001, *116*(4), pp. 1233-60.

Gul, Faruk. "A Theory of Disappointment Aversion." *Econometrica*, 1991, *59*(3), pp. 667-686.

Gul, Faruk and Pesendorfer, Wolfgang. "Temptation and Self-Control." *Econometrica*, November 2001, *69*(6), pp. 1403-35.

Harris, Richard J. "Dissonance or Sour Grapes? Post-decision Changes in Ratings and Choice Frequencies." *Journal of Personality and Social Psychology*, 1969, *11*(4), pp. 334-44.

Heath, Chip; Larrick, Richard P. and Wu, George. "Goals as reference points." *Cognitive Psychology. Special Issue: Belief and decision: The continuing legacy of Amos Tversky*, 1999, *38*(1), pp. 79-109.

Heidhues, Paul and Kőszegi, Botond. "Loss Aversion, Price Stability, and Sales," Working Paper, February 2004.

Higgins, Tory E. "Self-discrepancy: A Theory Relating Self and Affect." Psychological Review, July 1987, 94(3), pp. 319-40.

Higgins, Tory E. "Continuities and Discontinuities in Self-Regulatory and Self-Evaluative Processes: A Developmental Theory Relating Self and Affect." Journal of Personality, June 1989, 57(2), pp. 407-444.

Kahneman, Daniel. "Evaluation by Moments: Past and Future," in Daniel Kahneman and Amos Tversky, eds., *Choices, Values, and Frames*. New York: Russell Sage Foundation and Cambridge University Press, 2000, pp. 693-708.

Kahneman, Daniel; Knetsch, Jack L. and Thaler, Richard H. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy*, December 1990, *98*(6), pp. 1325-48.

Kahneman, Daniel; Knetsch, Jack L. and Thaler, Richard H. "The Endowment Effect, Loss Aversion, and Status Quo Bias: Anomalies." *Journal of Economic Perspectives*, Winter 1991, *5*(1), pp. 193-206.

Kahneman, D. and Lovallo, D. (1993), "Timid choices and bold forecasts. A cognitive perspective on risk taking." Management Science, 39, 17-31.

Kahneman, Daniel and Novemsky, Nathan. "Loss Aversion in Riskless and Risky Transactions: Choices, Exchanges and Gambles." Working Paper, Princeton University, 2002.

Kahneman, Daniel and Tversky, Amos. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, March 1979, *47*(2), pp. 263-91.

Kőszegi, Botond. "Anticipation in Observable Behavior." Working Paper, 2003. Previously: "Who Has Anticipatory Feelings?", 2001.

List, John. "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics*, February 2003, *118*(1), pp. 41-71.

Loewenstein, George; O'Donoghue, Ted and Rabin, Matthew. "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics*, November 2003, *118*(4), pp. 1209-1248.

Machina, Mark J. "Expected Utility Analysis without the Independence Axiom." *Econometrica*, March 1982, *50*(2), pp. 277-323.

Masatlioglu, Y. and Ok, Efe. "Rational Choice with Status-Quo Bias." Working Paper, New York University, 2002.

Mellers, Barbara; Schwartz, Alan and Ritov, Ilana. "Emotion-Based Choice." *Journal of Experimental Psychology: General*, September 1999, *128*(3), pp. 332-45.

Neumark, David and Postlewaite, Andrew. "Relative Income Concerns and the Rise in Married Women's Employment." *Journal of Public Economics*, October 1998, *70*(1), pp. 157-83.

Putler, Daniel S. "Incorporating Reference Price Effects into a Theory of Consumer Choice." *Marketing Science*, Summer 1992, *11*(3), pp. 287-309.

Rabin, Matthew. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica*, September 2000a, *68*(5), pp. 1281-92.

Rabin, Matthew. "Diminishing Marginal Utility of Wealth Cannot Explain Risk Aversion," in Daniel Kahneman and Amos Tversky, eds., *Choices, Values, and Frames.* New York: Cambridge University Press, 2000b, pp. 202-08.

Rabin, Matthew and Thaler, Richard H. "Risk Aversion." *Journal of Economic Perspectives*, Winter 2001, *15*(1), pp. 219-32.

Read, Daniel, Loewenstein, George, and Rabin, Matthew, "Choice Bracketing," Journal of Risk and Uncertainty 19 (1999), pp. 171-197.

Ryder, Harl E. and Heal, Geoffrey M. "Optimum Growth with Intertemporally Dependent Preferences." *Review of Economic Studies*, January 1973, *40*(1), pp. 1-33.

Sagi, Jacob. "Anchored Preference Relations." Working Paper, University of California, Berkeley, 2002.

Strahilevitz, Michal A. and Loewenstein, George. “The Effect of Ownership History on the Valuation of Objects.” *The Journal of Consumer Research*, December 1998, *25*(3), pp. 276-89.

Sugden, Robert. “Reference-Dependent Subjective Expected Utility.” *Journal of Economic Theory*, August 2003, *111*(2), pp. 172-191.

Tversky, Amos and Kahneman, Daniel. “Loss Aversion in Riskless Choice: A Reference-Dependent Model.” *Quarterly Journal of Economics*, November 1991, *106*(4), pp. 1039-61.