Reference-Dependent Risk Attitudes

By Botond Kőszegi and Matthew Rabin*

We use Kőszegi and Rabin's (2006) model of reference-dependent utility, and an extension of it that applies to decisions with delayed consequences, to study preferences over monetary risk. Because our theory equates the reference point with recent probabilistic beliefs about outcomes, it predicts specific ways in which the environment influences attitudes toward modest-scale risk. It replicates "classical" prospect theory—including the prediction of distaste for insuring losses—when exposure to risk is a surprise, but implies first-order risk aversion when a risk, and the possibility of insuring it, are anticipated. A prior expectation to take on risk decreases aversion to both the anticipated and additional risk. For large-scale risk, the model allows for standard "consumption utility" to dominate reference-dependent "gain-loss utility," generating nearly identical risk aversion across situations. (*JEL* D81)

Daniel Kahneman and Amos Tversky's (1979) prospect theory, and the literature building from it, provide theories of risk attitudes based on a few regularities. Most importantly, evaluation of an outcome is influenced by how it compares to a reference point, with people exhibiting both a significantly greater aversion to losses than appreciation of gains, and a diminishing sensitivity to changes in an outcome as it moves farther from the reference point. In addition, people weight the probability of a prospect nonlinearly, overweighting small probabilities and underweighting high probabilities.

The implications of prospect theory have been studied with several different specifications of the reference point, including the status quo, lagged status quo, and the mean of the chosen lottery. These various approaches explain many risk attitudes that are inconsistent with the classical diminishing-marginal-utility-of-wealth model, but they also generate mutually inconsistent predictions that to our knowledge have not been formally reconciled. Under a status quo specification, loss aversion predicts the substantial dislike of modest-scale risks involving both gains and losses that has been widely observed, and diminishing sensitivity predicts the risk lovingness in high-probability losses found by many researchers in the laboratory.[1] Under specifications based on the lagged status quo—such as in Richard H. Thaler and Eric J. Johnson (1990) and Francisco Gomes (2005)—diminishing sensitivity predicts the willingness to take unfavorable risks to regain the previous status quo. This "disposition effect," which has been observed by Terrance Odean (1998) for small investors and by David Genesove and Christopher Mayer (2001) for homeowners, is inconsistent with the substantial risk aversion predicted by a status quo model for gambles involving both gains

* Kőszegi: Department of Economics, University of California, Berkeley, 549 Evans Hall, 3880, Berkeley, CA 94720 (e-mail: botond@econ.berkeley.edu); Rabin: Department of Economics, University of California, Berkeley, 549 Evans Hall, 3880, Berkeley, CA 94720 (e-mail: rabin@econ.berkeley.edu). We are grateful to Paige Skiba and Justin Sydnor for research assistance, and to Erik Eyster, Vince Crawford, and three anonymous referees for detailed suggestions. We also thank seminar participants at Harvard University, London School of Economics, Stockholm School of Economics, and UC Berkeley for comments. Rabin thanks the National Science Foundation for financial support under grant SES-0518758, and Kőszegi thanks the Central European University for its hospitality while some of this work was completed.

[1] As has been pointed out by many researchers and formalized by Rabin (2000), nontrivial modest-scale risk aversion is calibrationally inconsistent with the classical diminishing-marginal-utility-of-wealth model. For instance, a person with $1 million in lifetime wealth who has CRRA utility and who rejects a "fifty-fifty lose $500 or gain $550" gamble would also turn down an equal-probability bet of losing $4,000 or gaining $100,000,000,000,000. While nobody would turn down the large-scale bet, most people would turn down the smaller gamble. Nicholas Barberis, Ming Huang, and Thaler (2006), for example, find that the majority of MBA students, financial advisors, and even very rich investors (with median financial wealth over $10 million) reject the $500–$550 bet.

and losses relative to the current status quo. And under specifications based on the chosen lottery's certainty equivalent—such as in the disappointment-aversion models of David E. Bell (1985), Graham Loomes and Robert Sugden (1986), and Faruk Gul (1991)—loss aversion implies substantial aversion to any risk. This strong risk aversion, which is apparent in consumers' choice of low insurance deductibles and purchase of extremely expensive extended warranties and automobile service contracts, is inconsistent with the risk lovingness in losses found in the lab and in the case of the disposition effect.

This paper uses the model from Kőszegi and Rabin (2006) and an extension to study monetary risk, unifying the seemingly different risk attitudes noted above as manifestations of the same preferences in different domains, and making novel predictions about behavior in situations not studied in the related literature. Our model (a) combines the reference-dependent "gain-loss utility" with standard "consumption utility"; (b) bases the reference point to which outcomes are compared on endogenously determined lagged beliefs; and, to incorporate probabilistic beliefs, (c) allows for stochastic reference points. Because of feature (a), our theory is consistent with aversion to all large-scale risk as predicted by classical expected-utility theory. Because of feature (b), it predicts both risk lovingness in response to *surprise* modest-scale losses, and—since anticipated premium payments do not generate sensations of loss while bad outcomes in uncertain situations do—first-order risk aversion when a risk and the possibility to insure it are *expected*. Hence, our theory matches both status quo prospect theory and disappointment aversion in domains where these models have been applied, and more generally provides comparative-statics predictions on the extent of risk taking as a function of the environment. Because of features (b) and (c), our theory predicts that the prior expectation of risk, even if it can now be avoided, decreases risk aversion. Unlike all the theories above, therefore, it predicts less risk aversion when deciding whether to remove expected risk than when deciding whether to take on that risk.

For a wealth level w and reference wealth level r, Section I specifies a person's utility as $u(w|r) \equiv m(w) + \mu(m(w) - m(r))$. The reference-independent "consumption utility," $m(w)$, corresponds to the classical notion of outcome-based utility. Gain-loss utility, $\mu(m(w) - m(r))$, depends on the difference between the consumption utility of the outcome and of the reference level, with the shape of μ corresponding to the loss aversion and diminishing sensitivity of prospect theory. Some of our results are established by assuming only what is commonly taken to be the stronger of these two forces, loss aversion.

We assume that the reference point relative to which a person evaluates an outcome is her *recent beliefs* about that outcome. An employee who had expected a \$50,000 salary will assess a salary of \$40,000 as a loss, and a taxpayer who had expected to pay \$30,000 in taxes will treat a \$20,000 tax bill as a gain. Because a person may be uncertain about outcomes, our theory allows for the reference point to be a distribution $G(\cdot)$, with a wealth outcome w then evaluated with "mixed feelings" as the average of the different assessments $u(w|r)$ generated by the r possible under $G(\cdot)$. For simplicity, we abstract from nonlinear decision weights: given a (stochastic or deterministic) reference point, a stochastic wealth outcome is evaluated according to its expected reference-dependent utility.

Our model of how utility depends on beliefs could be combined with any theory of how these beliefs are formed. As an imperfect but at the same time disciplined and largely realistic first pass, we assume that a person correctly predicts her probabilistic environment and her own behavior in that environment, so that her beliefs fully reflect the true probability distribution of outcomes. We begin in Section II by considering "surprise" (low-probability) decisions, modelled in extreme form as situations where expectations are given exogenously to the actual choice set. To illustrate implications for modest-scale risk, where consumption utility is approximately linear, consider a person's decision on whether to pay \$55 to insure a 50 percent chance of having to pay \$100. If she had expected to retain the status quo of \$0, our model makes the same prediction as prospect theory: because of diminishing sensitivity, she does not wish to insure the risk. If she had expected to pay \$55 for insurance, however, paying that amount generates no gain or loss, while taking the gamble exposes her to a fifty-fifty chance of losing \$45 or gaining \$55. With a conventional estimate of two-to-one loss aversion, she strongly dislikes this gamble and buys the insurance. Yet if a person had been expecting risk to start with, paying $0 instead of $100 can decrease expected losses, and paying $100 might just decrease expected gains, so the gamble is less aversive. When the ex ante expected risk is the gamble itself, this decreased risk aversion can be interpreted as an endowment effect for risk. When the ex ante expected uncertainty is very large, $100 cannot much change the extent to which money is evaluated as a loss rather than a gain, so the person is close to risk neutral.

In Sections III and IV, we study attitudes to anticipated risks. We identify two implications of our model: a person is more risk averse when she anticipates a risk and the possibility to insure it than when she does not—always displaying first-order risk aversion—and among such decisions regarding anticipated risk, she is more risk averse when she can commit to insure ahead of time.

When a decision is made shortly before the outcomes resulting from it occur, at that moment the reference point is fixed by past expectations, so that the decision maker maximizes expected utility taking the reference point as given. Being fully rational, therefore, she can expect behavior only if she is willing to follow it through, given a reference point determined by the expectation to do so. Formalizing this idea, in Section III we import from Kőszegi and Rabin (2006) the concept of an "unacclimating personal equilibrium" (UPE), defined as behavior where the stochastic outcome generated by utility-maximizing choices conditional on expectations coincides with expectations. Positing that a person can make any plans she knows she will follow through, our analysis assumes that she chooses her favorite UPE, the "preferred personal equilibrium" (PPE).

Applying PPE, we predict a very strong taste for planning and executing the purchase of small-scale insurance. The reason is a formalization and elaboration of some previous researchers' (e.g., Kahneman and Tversky 1984) psychological intuition that money given up in regular budgeted purchases is not a loss. In our model, a bad outcome of an *uncertain* lottery is evaluated as a loss, but a fully expected premium payment is not evaluated as a loss. Loss aversion therefore generates first-order risk aversion toward all insurable risks.

When a person makes a committed decision long before outcomes occur, she affects the reference point by her choice. For these situations, we introduce in Section IV the idea of a "choice-acclimating personal equilibrium" (CPE), defined as a decision that maximizes expected utility given that it determines both the reference lottery and the outcome lottery. Except that we specify the reference point as a lottery's full distribution rather than its certainty equivalent, this concept is similar to the disappointment-aversion models of Bell (1985), Loomes and Sugden (1986), and Gul (1991). Like PPE, CPE predicts that the decision maker strongly prefers to insure expected risks. But there is also an important difference. In some situations, a person would be better off with the reduced uncertainty of expecting to buy expensive insurance than not expecting to do so, but always choosing to avoid the expense of insurance at the last minute if not committed to buying it. Hence, the distaste for the risk manifests itself in behavior when the insurance decision is made up front, as in CPE, but not when the decision is made later, as in PPE. In such situations, environments where CPE is appropriate generate greater risk aversion and higher expected utility than environments where PPE applies.

The sensitivity of behavior to the economic environment described above applies only to modest-scale choices, where risk attitudes are necessarily dominated by the gain-loss component of preferences. In Section V, we investigate attitudes toward large-scale risk, where consumption utility cannot be assumed to be linear. We show that under reasonable conditions, the reference point has only a minor impact on how a person evaluates very large gambles. A person is therefore prone to exhibit risk aversion reflecting diminishing marginal utility of wealth independently of the environment.

Beyond helping to explain in a unified framework seemingly contradictory behavior, we hope the endogenous specification of the reference point helps make our model readily portable to many settings. To facilitate applications, in Appendix A we present an array of risk-characterization concepts and results. In Section VI, we conclude the paper by discussing some of the shortcomings of our model, emphasizing especially its failure to capture important ways that reference-dependent risk attitudes reflect failures of full rationality.

I. Reference-Dependent Utility

In this section we present the one-dimensional version of the utility function in Kőszegi and Rabin (2006). As is standard for models of risky choice outside of full-fledged life-cycle consumption models in macroeconomics, our theory takes as a primitive the choice set of gambles a person focuses on, in isolation from other risks and choices she faces.

For a riskless wealth outcome $w \in \mathbb{R}$ and riskless reference level of wealth $r \in \mathbb{R}$, utility is given by $u(w|r) \equiv m(w) + \mu(m(w) - m(r))$.[2] The term $m(w)$ is intrinsic "consumption utility" usually assumed relevant in economics, and the term $\mu(m(w) - m(r))$ is the reference-dependent gain-loss utility. Our model assumes that how a person feels about gaining or losing relative to a reference point depends on the changes in consumption utility associated with such gains or losses. This separation and interdependence of economic and psychological payoffs is analogous to assumptions made previously by Bell (1985), Loomes and Sugden (1986), and Veronika Köbberling and Peter P. Wakker (2005).

To accommodate our assumption below that the reference point is beliefs about outcomes, we allow for the reference point to be a probability measure G over $\mathbb{R}$:

$$U(w|G) = \int u(w|r)\ dG(r). \tag{1}$$

This formulation captures the notion that the evaluation of a wealth outcome is based on comparing it to all possibilities in the support of the reference lottery. For example, if the reference lottery is a gamble between \$0 and \$100, an outcome of \$50 evokes a mixture of two feelings, a gain relative to \$0 and a loss relative to \$100.[3]

When w is drawn according to the probability measure F, utility is given by

$$U(F|G) = \iint u(w|r)\ dG(r)dF(w). \tag{2}$$

For simplicity and contrary to Kahneman and Tversky (1979) and its extensions, we assume that preferences are linear in probabilities. This means that our model will get some predictions—e.g. regarding insurance of low-probability losses—wrong.

Our utility function is closely related to that of Sugden (2003). In his model, outcome lotteries are compared to reference lotteries state by state, capturing a form of state-contingent disappointment missing from our theory.[4] Another alternative to our formulation, pursued by Bell (1985), Loomes and Sugden (1986), Gul (1991), and Jonathan Shalev (2000), is to collapse the reference lottery into some type of certainty equivalent.[5] With such a specification, two

[2] Like many models, this paper assumes preferences are over monetary wealth rather than consumption. Strictly speaking, this means our model corresponds to a single-period setting. While we have not verified how our results extend to a multiperiod consumption model, the shortcut of using wealth may be appropriate even in that case, both because people often experience sensations of gain and loss directly from wealth changes, and because wealth is a summary statistic for consumption and hence may generate similar gain-loss sensations. As a person's wealth increases, for example, her anticipation of increased consumption throughout the future is likely to generate a sense of gain similar to that presumed in our model. A developing body of research on dynamic models of reference-dependent utility, such as Rebecca Stone (2005) and Kőszegi and Rabin (2007), may help extend and explore the robustness of the results we develop in this paper.

[3] More than saying a person separately compares an outcome to all components of the reference lottery, our formulation of $\mu(\cdot)$ below implies that losses relative to a stochastic reference point count more than gains, so that the \$50 above yields negative gain-loss utility. An alternative specification is one where the relief of avoiding the \$0 outcome outweighs the disappointment of not getting the \$100 outcome. This alternative seems difficult to reconcile with loss aversion relative to riskless reference points. It would also seem to imply that people will seek to endow themselves with risks because the chance for a pleasant relief from getting better outcomes outweighs the potential disappointment from getting bad outcomes.

[4] While this state-contingency seems in some cases to be more realistic than our approach, and we do not know the extent to which the two models can be reconciled within a broader framework, our prediction of a state-independent disappointment when receiving worse-than-expected outcomes seems pervasively realistic, and is missing from Sugden's approach. In addition, Sugden's model (unlike ours) makes the implausible prediction that a person becomes very risk loving when given the option to replace the reference lottery with a riskless amount.

[5] There is some suggestive evidence of mixed feelings when there are multiple counterfactuals relative to which outcomes can be evaluated. For instance, Jeff T. Larsen et al. (2004) find that when a subject receives \$5 from a lottery that could have paid \$5 or \$9, she has both positive and negative emotions—presumably from winning \$5 and not

reference lotteries that have the same certainty equivalent generate the same risk preferences. This is inconsistent with our theory's prediction that a person is more inclined to accept a risk if she had been expecting risk, a prediction that seems broadly correct based on the little available evidence. But as we discuss below, the main difference between all these models and ours is in the specification of the reference point.

We assume 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 $y > x \geq 0$, 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. $\mu'_-(0)/\mu'_+(0) \equiv \lambda > 1$, where $\mu'_+(0) \equiv \lim_{x\to 0} \mu'(|x|)$ and $\mu'_-(0) \equiv \lim_{x\to 0} \mu'(-|x|)$.

Properties *A0–A4*, first stated by David Bowman, Deborah Minehart, and Rabin (1999), correspond to Kahneman and Tversky's (1979) explicit or implicit assumptions about their "value function" defined on $w - r$. Loss aversion is captured by *A2* for large stakes and *A4* for small stakes, and diminishing sensitivity is captured by *A3*. While the inequalities in *A3* are most realistically considered strict, to characterize the implications of loss aversion without diminishing sensitivity as a force on behavior, we define a subcase of *A3*:

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

When we apply *A3′* below, we will parameterize μ as $\mu'_+(0) = \eta$ and $\mu'_-(0) = \lambda\eta > \eta$, so that η can be interpreted as the weight attached to gain-loss utility.

To determine behavior, the utility function introduced above needs to be combined with a theory of reference-point determination. As a disciplined and largely realistic first pass, we assume that a person's reference point is the *rational expectations* about the relevant outcome she held between the time she first focused on the outcome and shortly before it occurs. For example, if an employee had been expecting a salary of \$100,000, she would assess a salary of \$90,000, not as a large gain relative to her status quo wealth, but as a loss relative to her expectations of wealth.[6] As we explain in detail in Kőszegi and Rabin (2006), our primary motivation for equating the reference point with expectations is that this assumption helps unify and reconcile existing intuitions and discussions. We assume *rational* expectations both to maintain modelling discipline (much like in other rational-expectations theories), and because we feel in most situations people have some ability to predict their own environment and behavior. Unfortunately, relatively little evidence on the determinants of reference points currently exists. Some existing evidence does, however, provide empirical support for expectations as a component of the reference point. In analyzing play in the huge-stakes game show "Deal or No Deal," for example, Thierry Post et al. (forthcoming) find evidence that past expectations affect behavior. In the game, a contestant "owns" a suitcase with a randomly determined prize. Gradually, the contestant learns information about the prize in her bag (by opening other bags and learning what is not in her bag). At each stage, a "bank" offers a riskless amount of money to replace the amount in the bag. A contestant's acceptance or rejection of the offer is an indication of her risk aversion. A key finding is that contestants reject better offers when they have received bad news in the last few rounds, suggesting that they are less risk averse in these contingencies.[7]

winning \$9, respectively. Losing \$5 when it is also possible to lose \$9 evokes similar mixed feelings.

[6] Our theory posits that preferences depend on *lagged* expectations, rather than on expectations contemporaneous with the time of consumption. This does not assume that beliefs are slow to adjust to new information or that people are unaware of the choices they have just made—but that preferences do not instantaneously change when beliefs do. When somebody finds out five minutes ahead of time that she will for sure not receive a long-expected \$100, she presumably immediately adjusts her expectations to the new situation, but five minutes later she will still assess not getting the money as a loss.

[7] For further examples of evidence of expectations-based counterfactuals affecting reactions to outcomes,

Many researchers have noted over the years that the reference point may to some extent be influenced by expectations. But most previous formal models either equate the reference point with the status quo, or leave it unspecified, and none explicitly equates it to recent beliefs about outcomes. To our knowledge, the disappointment-aversion models by Bell (1985), Loomes and Sugden (1986), and Gul (1991), come closest to saying that the reference point is recent expectations.[8] But because these models assume the reference point is formed after choice, they treat surprise situations differently from our theory, predicting in particular first-order risk aversion for surprise losses. Our model can be thought of in part as unifying prospect theory and disappointment-aversion theory in one framework, while also proposing a solution concept (PPE) for situations where choices are anticipated but not committed to in advance. We are not aware of any model that attempts such a unification.

Although our model is in the tradition of most of economics in that it posits a utility function with certain properties, it differs from much of the foundational literature on choice under uncertainty in that it does not derive the utility function from axioms capturing those properties. It also differs from most economic theories in making explicit how utility depends on a mental state—beliefs—that is not directly observable in choice behavior. Obviously, neither of these features means that our model does not have observable or falsifiable implications. Indeed, in Appendix B we show how to extract the full utility function ($m(\cdot)$ and $\mu(\cdot)$) from behavior in a limited set of decision problems, and because our model provides a complete mapping from decision problems to possible choices, this completely ties down predictions for all other decision problems. And the many propositions in the paper derive general restrictions that our model implies for observed behavior.

see Victoria Husted Medvec, Scott F. Madey, and Thomas Gilovich (1995), Barbara Mellers, Alan Schwartz, and Ilana Ritov (1999), and Hans C. Breiter et al. (2001).

[8] In addition, the notion of "loss-aversion equilibrium" that Shalev (2000) proposed for multiplayer games can be interpreted as saying that each player's reference point is expectations. As we discuss below, rewriting Shalev's model using our specification of stochastic reference points and applying it to individual decision making corresponds to the UPE solution concept.

II. Risk Attitudes in Surprise Situations

In the next three sections, we investigate the decision maker's attitudes toward modest-scale risk, such as \$100 or \$1,000, where consumption utility can be taken to be approximately linear—and where we therefore derive formal results under the assumption that $m(w) = w$.[9] In Section V, we return to an exploration of large-scale risk, where risk preferences can be substantially influenced by diminishing marginal utility of wealth. We organize our results on modest-scale risk into three sections according to the expectational environment the decision maker faces, considering in turn "surprise," PPE, and CPE situations. But a number of themes link the three sections. Propositions 1, 2, 5, and 6 identify common ways in which the decision maker becomes less risk averse if she had been expecting, or is now facing, more risk. Proposition 3 shows that a person is first-order risk averse when she anticipates a risk as well as the possibility to insure it. And Propositions 4, 7, and 8 show that the more a risk can be prepared for, the greater is the risk aversion displayed in behavior: the person's risk aversion is greater when she expects an insurance decision than when she does not, and even greater when she can commit early to purchasing insurance.

This section begins the analysis with risk-taking behavior when the reference point is fixed, considering both deterministic and stochastic reference points. The analysis is the limiting case of UPE/PPE behavior when the decision maker finds herself in an ex ante low-probability situation, so that she has fixed expectations formed essentially independently of the relevant choice set.

[9] If $m(\cdot)$ were not linear (but remained differentiable), some of our results would survive unchanged, and the others could be modified by restating them in terms of expected consumption utilities instead of expected values. But because this would complicate our statements, and because consumption utility is so close to linear for modest stakes, we assume $m(\cdot)$ is linear. To give a sense of the calibrational appropriateness of this approximation, note that even for a person who has a low \$1 million in lifetime wealth and a very high consumption-utility coefficient of relative risk aversion of 10, winning or losing \$1,000 (a difference of \$2,000 in wealth) changes marginal consumption utility by only 1.8 percent.

Proposition 2 of Kőszegi and Rabin (2006) shows that when the decision maker expects to keep the status quo, her behavior is identical to that predicted by prospect theory modified to equate decision weights with probabilities. Hence, our model is consistent with much of the evidence motivating status quo prospect theory. It can also be used to interpret the "disposition effect" found by Odean (1998) for stocks and Genesove and Mayer (2001) for houses—whereby people appear disproportionately reluctant to sell an asset for less than they paid.[10] The intuition commonly invoked for the disposition effect, formalized by Gomes (2005) in a version of prospect theory based on the lagged status quo, is that because the purchase price operates as a reference point for the selling price, people will be risk-seeking in waiting for a price to recover before selling.[11] While our model cannot explain all aspects of the disposition effect, by basing the reference point on the expected resale price, it does help to determine when and how the effect is likely to be observed. Because home and stock owners usually expect to make money, they will be risk loving when these investments unexpectedly lose money. But when a person foresees a good chance of losing some money—for example, when investing in a highly risky stock—she will be less willing to take chances to break even. And if she expects an investment—such as a house in a booming market or inventory a merchant expects to resell at a large margin—to make a large positive return, she may even be reluctant to sell at prices insufficiently above the purchase price.

While our model only replicates or qualifies classical prospect theory in the settings above, it makes a set of novel predictions regarding the effect of prior uncertainty on behavior, identifying senses in which the expectation of risk decreases aversion to the expected as well as additional risks. To state these results, we use $H + H'$ to denote the distribution of the sum of independent draws from the distributions H and H'. (Thus, $(H + H')(z) = \int H(z - s)\, dH'(s)$.) When it creates no confusion, a real number will denote both a deterministic wealth level and the lottery that assigns probability 1 to that amount of wealth. Proposition 1 says that under $A3'$, a person is no more willing to accept a given lottery if it is added to a riskless reference point than if it is added to a lottery and/or evaluated relative to a risky reference point.

PROPOSITION 1: *Suppose $m(\cdot)$ is linear and $\mu(\cdot)$ satisfies $A3'$. For any lotteries F, G, and H and constant w, if $U(w + F|w) \geq U(w|w)$, then $U(H + F|G) \geq U(H|G)$.*

Since $m(\cdot)$ is linear and A3′ is satisfied, a small change in an outcome is evaluated solely according to the previously expected probabilities of getting higher and lower outcomes, and not according to the distance from those higher and lower outcomes. Now when F is added to a riskless reference point w, positive outcomes of F are assessed as pure gains, and negative outcomes of F are assessed as pure losses. But when F is instead added to a lottery H and is evaluated relative to a lottery G, positive outcomes of F partially eliminate losses suffered from H relative to G, and are hence evaluated more favorably than pure gains; and negative outcomes of F in part merely eliminate gains from H relative to G, and are hence evaluated less unfavorably than pure losses. For both these reasons, the decision maker is more willing to accept F.

An important implication of Proposition 1—obtained by setting $H = w$ and $G = F$—is a type of endowment effect for risk: a person is less risk averse in eliminating a risk she expected to face than in taking on the same risk if she did not expect it. This prediction of our model contrasts with previous theories of reference-dependent utility with which we are familiar. While little evidence on the issue seems to be available, choice experiments by Jack L. Knetsch and J. A. Sinden (1984) and evidence on hypothetical choices in Michael H. Birnbaum et al. (1992) do

[10] The disposition effect is closely related to the "break-even effect" coined by Thaler and Johnson (1990) for monetary gambles. They predicted that following losses, gambles that offer a chance to break even become especially attractive.

[11] Barberis and Wei Xiong (2006) show that prospect theory based on the lagged status quo does not necessarily imply an increased risk lovingness after losses. Intuitively, for risks a person takes on voluntarily, losses are typically smaller than gains. Hence, a person is typically closer to the reference point after a loss than after a gain, and due to loss aversion this can mean greater aversion to substantial amounts of risk. In the Odean (1998) and Genesove and Mayer (2001) studies, however, individuals can take more incremental risk—waiting more or less exactly until the price returns to the reference point.

indicate that people tend to be less risk averse when selling a lottery they are endowed with than when buying the lottery.

A second way in which expecting risk decreases risk aversion is that a person is approximately risk neutral in accepting a lottery that is "small" relative to the reference lottery.

PROPOSITION 2: *Suppose $m(\cdot)$ is linear. For any lottery F with positive expected value:*

(*i*) *There exist A, $\varepsilon > 0$ such that if the lottery G satisfies $Pr_G[r \in (k - A, k + A)] < \varepsilon$ for all constants k, then $U(H + F|G) > U(H|G) > U(H - F|G)$ for any lottery H.*

(*ii*) *For any continuously distributed lottery G, there is a $\bar{t} > 0$ such that for any $t \in (0, \bar{t}\,]$ and any lottery H, $U(H + t \cdot F|G) > U(H|G) > U(H - t \cdot F|G)$.*

The proposition identifies attitudes toward a given lottery *F* with positive expected value, showing ways in which the decision maker is risk neutral when evaluating multiples of *F*. Part 1 says that if the reference lottery is sufficiently widely distributed, the decision maker will take *F* and reject $-F$. Part 2 says that fixing a continuously distributed reference lottery, she will take a sufficiently small multiple of *F* and reject the same multiple of $-F$. Intuitively, if the extra lottery is small relative to the reference lottery *G*, it is unlikely to turn a gain or loss relative to *G* into the opposite, and there is little diminishing sensitivity over its range. Hence, neither loss aversion nor diminishing sensitivity plays a major role in its evaluation.

Part 2 of Proposition 2 is related to Proposition 1 in Barberis, Huang, and Thaler (2006) showing that a loss-averse (in their terminology "first-order risk-averse") decision maker is only second-order risk averse if she faces background risk. In addition, Proposition 2 demonstrates that even if a person now faces no background risk (*H* is riskless), the mere prior *expectation* of risk makes her second-order risk averse.

Crucially for a number of later results, Propositions 1 and 2 do *not* imply that a person is unbothered by risk she expects or faces, even if it does make her less risk averse. In fact, the same force that decreases risk aversion also decreases expected utility: when the decision maker had been expecting risk or is already facing risk, she is already exposed to stochastic utility-decreasing sensations of loss, so taking additional risk does not add so much exposure to losses.

To illustrate immediate and later points, Table 1 shows some of our model's implications in a parameterized example with consumption utility $m(w) = 10{,}000 \ln(w)$, and gain-loss utility $\mu(x) = \sqrt{x}$ for $x \geq 0$ and $\mu(x) = -3\sqrt{-x}$ for $x \leq 0$.[12] We consider three fifty-fifty gambles of different scales. Rows correspond to environments in which the gambles might be evaluated. For each gamble, the first column identifies the premium the decision maker is willing to pay for insurance against the risk in each environment, calculated as the difference between the gamble's mean and its certainty equivalent. The second column gives the coefficient of relative risk aversion $\tilde{\rho}$ that would be inferred from the gamble's certainty equivalent if one assumed reference-independent CRRA utility.

Gamble I is a small-stakes gamble that pays either \$1,000,100 or \$999,900. The first panel of the table considers surprise situations with various (exogenous) expectations. The decision maker displays extreme risk lovingness if the gamble is a small loss relative to expectations, and extreme risk aversion if it is a small gain or involves both a loss and a gain. Although loss aversion is a stronger force than diminishing sensitivity, the decision maker displays more risk aversion for gains than for mixed gambles because in the latter case she evaluates the certainty equivalent as a loss.[13] And as Propositions 1 and 2 predict, the prior expectation of risk decreases risk aversion, with expected risk on the order of \$1,000 or \$10,000 already lowering

[12] Note that, unless $A3'$ holds, our model is not invariant to affine transformations of $m(\cdot)$, so the appropriate specification of $m(\cdot)$ and $\mu(\cdot)$ involves a substantive assumption about their relative scaling. This scaling amounts to an assumption about the speed of diminishing sensitivity in gain-loss utility. Indeed, the choice to specify our example with $m(w) = 10{,}000\ln(w)$ is designed to get gain-loss utility to dominate for an appropriate range of small stakes, and consumption utility to dominate for an appropriate range of large stakes.

[13] Indeed, while not shown in the table, the riskless reference point that makes the person indifferent between the reference point and the gamble—a situation where the insurance premium is not evaluated as a loss—corresponds to a much higher, \$48 premium for insurance.

TABLE 1—ATTITUDES TOWARD THREE 50–50 GAMBLES

	Gamble I $999,900/$1,000,100		Gamble II $990,000/$1,010,000		Gamble III $500,000/$1,500,000	
	WTP	$\tilde{\rho}$	WTP	$\tilde{\rho}$	WTP	$\tilde{\rho}$
Surprise with different (fixed) expectations						
$500,000	$0.01	1.01	$50	1.01	$135,840	1.02
$990,000	0.02	3.4	329	6.5	132,477	0.99
$999,900	19.2	3,936	469	9.4	132,248	0.99
$1,000,000	9.17	1,844	432	8.6	132,246	0.99
$1,000,100	−33.1	−7,171	397	7.9	132,244	0.99
$1,010,000	−<0.01	−5.6	−712	−14	132,028	0.99
$1,500,000	<0.01	0.97	49	0.97	128,429	0.96
U[$999,900, $1,000, 100]	3.66	733	433	8.7	132,246	0.99
U[$999,000, $1,001,000]	0.27	52	507	10.2	132,246	0.99
U[$990,000, $1, 010, 000]	0.02	3.5	275	5.5	132,248	0.99
PPE with different amounts of background risk						
none	$48.0	11,389	$757	15	$138,502	1.04
U[−$100, $100]	8.54	1,717	779	14	138,467	1.04
U[−$1,000, $1,000]	0.56	111	656	13	138,394	1.03
U[−$10,000, $10,000]	0.03	6.0	259	4.7	138,176	1.03
CPE with different amounts of background risk						
none	$71.0	23,348	$757	15	$138,502	1.04
U[−$100, $100]	28.5	6,041	779	14	138,467	1.04
U[−$1,000, $1,000]	1.84	368	656	13	138,394	1.03
U[−$10,000, $10,000]	0.07	14	362	7.2	138,176	1.03

Notes: The table assumes a consumption utility of $m(w) = 10,000 \ln(w)$ and a gain-loss utility of $\mu(x) = \sqrt{x}$ for $x \geq 0$ and $\mu(x) = -3\sqrt{-x}$ for $x < 0$. For each gamble and environment, we calculated the gamble's certainty equivalent in that environment. WTP is the difference between the gamble's mean ($1,000,000 in each case) and the certainty equivalent. This is the premium the decision maker is willing to pay for insurance. $\tilde{\rho}$ is the coefficient of relative risk aversion that would be inferred from the gamble's certainty equivalent if one assumed reference-independent CRRA utility. $U[x, y]$ refers to a discrete uniform distribution between x and y, with atoms on multiples of $50.

the decision maker's willingness to pay for insurance to essentially zero.

III. UPE and PPE Risk Attitudes

While the previous section considered risk attitudes in surprise situations, our primary interest is in behavior when the decision maker correctly predicts the choice set she faces. From the psychological hypothesis that the reference point for evaluating outcomes is equal to lagged rational beliefs about those outcomes, we develop two reduced-form models that differ in when the decision maker makes a committed choice. In this section, we analyze her behavior in one extreme possibility, when she anticipates the decision she faces but cannot commit to a choice until shortly before the outcome. Although our model is ambiguous as to the interpretation of "shortly before," insurance choices on short-term rentals such as cars or skis probably best correspond to such a case. The resulting model is equivalent to the one-dimensional version of the model in Kőszegi and Rabin (2006).

Suppose the decision maker has probabilistic beliefs over possible compact choice sets described by $\{D_1, 1-q; D_2, q\}$, where choice sets $D_1, D_2 \subset \Delta(\mathbb{R})$ occur with probabilities $1-q$ and q, respectively. All of our results in the next two sections are for the case $q = 0$, so that the decision maker knows the single choice set she will face. But we keep our definition a little more general because—as we explain below—this allows us to capture the kinds of surprise situations discussed in the previous section as limiting cases of low-probability choice sets.

Since the person makes her decision shortly before the outcome resulting from it, at that time the beliefs determining the reference point are past and hence unchangeable. This means that she maximizes utility taking the reference point as given, so that she can rationally expect to follow a plan of behavior only if she is willing to follow it given a reference point generated by the expectation to do so.

DEFINITION 1: *A selection* $F_1 \in D_1, F_2 \in D_2$ *is an* unacclimating personal equilibrium *(UPE) if for each* $l \in 1, 2$ *and any* $F_l' \in D_l$, $U(F_l|(1-q)F_1 + qF_2) \geq U(F_l'|(1-q)F_1 + qF_2)$.[14]

If the person expects to choose F_1 and F_2 from choice sets D_1 and D_2, respectively, then given her expectations over possible choice sets, she expects the distribution of outcomes $(1-q)F_1 + qF_2$. Definition 1 states that with those expectations as her reference point, she should indeed be willing to choose F_1 and F_2 from choice sets D_1 and D_2.

UPE is closely related to the notion of "loss-aversion equilibrium" that Shalev (2000) defined for multiplayer games as a Nash equilibrium fixing each player's reference point, where the reference point is equal to the player's (implicitly defined) reference-dependent expected utility. Although Shalev does not himself pursue this direction, reformulating his notion of loss-aversion equilibrium using our utility function and applying it to individual decision making corresponds to UPE.

In the decision between a fifty-fifty chance of having to pay \$100 out of current wealth w and buying insurance for \$55, when is choosing the lottery a UPE? If the lottery is the reference point, the following inequality indicates when it is preferred to paying \$55:

$$\text{(3)} \quad \left[\frac{1}{2}(w-100) + \frac{1}{2}w\right] + \left[\frac{1}{4}\mu(100) + \frac{1}{4}\mu(-100)\right] \geq [w-55] + \left[\frac{1}{2}\mu(45) + \frac{1}{2}\mu(-55)\right].$$

[14] Because each of our solution concepts is an example of personal equilibrium as first defined in Kőszegi (2005), Theorem 1 of that paper implies that when $(1-q)D_1 + qD_2$ is convex and compact, UPE, as well as PPE and CPE, exist.

There can be multiple UPE in a given situation—there can be multiple self-fulfilling expectations—and generically different UPE yield different expected utilities. But the person's expectations are based on her own plans on what to choose once the time comes. It seems likely, therefore, that she will choose the best plan she knows she will follow through on.

DEFINITION 2: *A selection* $F_1 \in D_1, F_2 \in D_2$ *is a* preferred personal equilibrium *(PPE) if it is a UPE, and* $U((1-q)F_l + qF_2|(1-q)F_1 + qF_2) \geq U((1-q)F_1' + qF_2'|(1-q)F_1' + qF_2')$ *for all UPE selections* $F_l' \in D_1, F_2' \in D_2$.

A major feature of UPE and PPE is the constraint that choice must be optimal *given* expectations at the time. This means that the decision maker does not internalize the effect of her choice on expectations, so—as we will illustrate—she often does not maximize ex ante expected utility among the choices available to her.

To begin our analysis of UPE and PPE behavior, we note that the results in Section II on expectations fixed independently of the person's choice set D_1 can be thought of as applying UPE or PPE to situations where that choice set is a surprise: if she had been expecting to face choice set D_2 with near certainty and D_1 with very small probability, and to choose $F \in D_2$ (perhaps because $D_2 = \{F\}$, so that she thought she would have no choice), her reference point would be approximately F independently of D_1 or what she had been expecting to choose from D_1.

To study decisions for choice situations that are anticipated, we first establish that a person has a strong preference to insure modest-scale risks—she is first-order risk averse—and then show that expecting risk at the start decreases her aversion to additional risk.

PROPOSITION 3: *Suppose* $m(\cdot)$ *is linear. For any* $w \in \mathbb{R}$ *and mean-zero lottery* $F \neq 0$ *with bounded support, there exist* $\bar{k}, \bar{t} > 0$ *such that for any positive* $t < \bar{t}, k < \bar{k}$, *the unique PPE with the choice set* $\{w, w + t(F+k)\}$ *is to choose* w.

Proposition 3 states that a person will choose a riskless w over a sufficiently small multiple of a better-than-fair but insufficiently attractive bet.[15] Intuitively, w is a UPE because when the decision maker expects it, loss aversion leads her to turn down the gamble. And w is a *PPE* because the gamble exposes the decision maker to a sense of loss from bad outcomes that outweighs the sense of gain from good outcomes, and hence yields lower expected utility than w.

In predicting attitudes toward insuring moderate and high-probability losses, status quo prospect theory says both that loss aversion does not play a role, and that diminishing sensitivity pushes people toward not insuring. When it comes to insuring *expected* losses, our model reverses this prediction: it says that loss aversion plays the crucial role and pushes people toward insuring. As a key force behind this prediction, our model captures an intuition regarding the difference between "costs" and "losses" that has sometimes been articulated (e.g., in Kahneman and Tversky 1984 and Nathan Novemsky and Kahneman 2005), but has not been formalized. In our model, beliefs about wealth take into account a planned premium payment, so that such a payment is *not* evaluated as a loss. By contrast, whether or not a person had expected to buy insurance, a bad realization of a stochastic lottery *is* evaluated as a loss. Because loss aversion therefore plays a central role in the decision of whether to insure, first-order risk aversion results. Indeed, some consumers' purchase of insurance against high-probability losses—such as extremely expensive automobile service contracts—seems consistent with this prediction and is in direct contrast to status quo prospect theory.[16]

For low-probability losses—such as those covered by extended warranties and low deductibles on homeowners' insurance—the overweighting of low probabilities in conventional status quo prospect theory can counteract diminishing sensitivity and create aversion to risk. But even for these decisions, the extent of risk aversion often seems quantitatively much stronger than predicted by prospect theory. A calibrational problem is immediately apparent in intuitive terms if we pose the deductible choices of homeowners analyzed by Justin Sydnor (2006) as gambles relative to the status quo. For the average consumer with a \$500 deductible, the premium for a \$1,000 deductible is about \$600, and the premium for a \$500 deductible is about \$100 higher. The probability of making a claim in any given year is 0.05. Hence, the choice between the two deductibles is equivalent to the choice between the gambles $(-600, 0.95;\ -1600, 0.05)$ and $(-700, 0.95;\ -1200, 0.05)$.[17] We conjecture that when asked in these terms, most individuals would choose the former gamble, and indeed Sydnor (2006) shows that typical parameterizations of prospect theory make the same prediction. Yet the majority of consumers in his dataset chose the latter gamble.

But the difference between anticipated and surprise situations extends beyond attitudes toward losses. As indicated in the second panel of Table 1, the certainty equivalent for Gamble I in our example is lower in a PPE situation with no background risk than in any of the surprise situations. Indeed, if a person is deciding between taking on a risk and fully insuring it, under $A3'$ she is at least as risk averse in PPE as in a surprise situation, whatever her expectations may have been in the latter case.

PROPOSITION 4: *Suppose $m(\cdot)$ is linear and $\mu(\cdot)$ satisfies A3′. If $w + F$ is a PPE in the choice set $\{w, w + F\}$, then for any lottery H, $U(w + F|H) > U(w|H)$.*

Because of the many ways it predicts that behavior in surprise situations differs from that in expected situations, our theory cautions against extrapolating from experimental results too casually. Insofar as most experimental subjects do not have a clear idea about the tasks they are going to face, their behavior does not correspond to what they would do in a similar but expected

[15] Proposition 11 in Appendix A identifies a precise condition of the attractiveness of a vanishingly small lottery that determines whether the decision maker accepts the lottery.

[16] An experiment by Antoni Bosch-Domènech and Joaquim Silvestre (2006) can also be interpreted as providing evidence for risk aversion in high-probability losses. Subjects received money for their performance in the first session of the experiment, and were asked to come back weeks later for a second session. Subjects were warned that they could lose money during the second session. In this session, a majority of students were risk averse for both high- and low-probability losses.

[17] Strictly speaking, the two situations are not equivalent if some claims are between \$500 and \$1,000. But if so, a high deductible is all the more attractive.

situation in real life. Specifically, Proposition 4 says that laboratory measurements of small-scale risk aversion will often underestimate risk aversion for expected small-scale risks.

We conclude this section with results on two ways in which expecting risk at the start decreases aversion to additional risk. Although our formal results identify the effect of facing a lottery G the decision maker *cannot* avoid, this can be the reduced-form representation of a situation where she could, but (because it is too attractive) in equilibrium *does not* avoid G. First, the decision maker is no less likely to accept a given lottery if she is already facing risk than if she is not.

PROPOSITION 5: *Suppose $m(\cdot)$ is linear and $\mu(\cdot)$ satisfies A3′. For any lotteries G and F and constant w, if choosing G is a UPE with the choice set $\{G, G+F\}$, then choosing w is a PPE with the choice set $\{w, w+F\}$.*

And as a corollary to Proposition 2, a person is neutral to risks that are small relative to the risk she is already expecting.

PROPOSITION 6: *Suppose $m(\cdot)$ is linear. For any lottery F with positive expected value:*

(i) There exist $A, \varepsilon > 0$ such that for any lottery G for which $Pr_G(r \in [k-A, k+A]) < \varepsilon$ for all constants k, the unique UPE with the decision set $\{G, G+F\}$ is to choose $G+F$, and the unique UPE with the decision set $\{G, G-F\}$ is to choose G.

(ii) For any continuously distributed lottery G, there is a $\bar{t} > 0$ such that for any $t \in (0, \bar{t}]$, the unique UPE with the decision set $\{G, G+t\cdot F\}$ is to choose $G+t\cdot F$, and the unique UPE with the decision set $\{G, G-t\cdot F\}$ is to choose G.

The second panel of Table 1 quantifies Propositions 5 and 6 for our parameterized example, showing that the decision maker approaches risk neutrality for Gamble I even for relatively limited amounts of background risk. An unavoidable risk on the order of a mere \$100 decreases the premium she is willing to pay for insurance from \$48 to \$9, and unavoidable risk on the order of \$1,000 makes her virtually risk neutral.

IV. CPE Risk Attitudes

Beyond the distinction between surprise and anticipated exposure to risky decisions, our underlying assumptions also imply differences in attitudes toward anticipated risks as a function of how far in advance decisions are committed to. We now analyze risk preferences regarding outcomes that are resolved long after all decisions are committed to, a situation that applies to most insurance choices. In this case, the expectations relative to which a decision's outcomes are evaluated are formed after—and therefore incorporate the implications of—the decision.

DEFINITION 3: *For any choice set D, $F \in D$ is a* choice-acclimating personal equilibrium *(CPE) if $U(F|F) \geq U(F'|F')$ for all $F' \in D$.*

If the decision maker makes the choice $F \in D$ today, this will determine her reference point by the time the relevant wealth outcome occurs. Thus, when evaluating her resulting expected utility, both the reference and outcome lotteries are equal to F. Note that CPE can naturally be extended to decisions where some uncertainty is resolved today. In this case, immediately resolved uncertainty gets absorbed into the reference point, so that the decision maker maximizes the expectation of $U(F|F)$ taken over the possible lotteries F that capture solely the remaining uncertainty. Hence, for instance, if somebody is confronted with a choice between getting \$10,000 for certain a month later and getting \$25,000 a month later if an immediately observed coin flip comes up heads, we predict she would evaluate her options solely in terms of expected consumption utility.

Our notion of CPE is related to the models of "disappointment aversion" of Bell (1985), Loomes and Sugden (1986), and Gul (1991), where outcomes are also evaluated relative to a reference lottery that is identical to the chosen lottery. As we have mentioned, however, these theories differ from ours in two important ways. They assume that a person evaluates outcomes relative to the reference lottery's certainty equivalent instead of the full distribution. And because they do not distinguish expectational environments at all, they do not share many of our model's predictions in other situations.

Returning to our example of choosing between a 50 percent chance of losing \$100 and insuring this risk for \$55, selecting the lottery is a CPE if

$$(4)\qquad \left[\frac{1}{2}(w-100)+\frac{1}{2}w\right] + \left[\frac{1}{4}\mu(100)+\frac{1}{4}\mu(-100)\right] \geq [w-55]+[0].$$

The difference between UPE and CPE is in the right-hand sides of inequalities (3) and (4), which capture the decision maker's expected utilities when deviating from the purported UPE and CPE, respectively. In UPE, the reference point does not adjust to the deviation, so paying \$55 is assessed partly as a loss of \$55 and partly as a gain of \$45. In CPE, the reference point does adjust to the deviation, so there is no sensation of gain or loss when paying the \$55.[18]

As with PPE, except in knife-edge cases there will be a unique CPE. But unlike in PPE, where the decision maker can choose her favorite plan only from those that she would follow through on, in CPE she can commit to her overall favorite lottery. Hence, there cannot be a divergence between behavior and welfare.

It bears emphasizing that UPE/PPE and CPE are not different theories of what outcomes people prefer. Indeed, in our example the expected utility from choosing the lottery (the left-hand side of inequality (3) or (4)) is independent of whether the choice is determined by UPE or CPE. Rather than reflecting different notions of reference-dependent utility, the two concepts are motivated by the same theory of preference, as manifested differently depending on whether the person can commit to her choice ahead of time. Of course, one weakness of our theory is that it does not specify the lag with which new beliefs are incorporated into the reference point, so that it may be unclear whether a person's committed decision is made sufficiently early that CPE rather than PPE is appropriate for modeling her behavior. In many situations, this will not be a major problem, as the appropriate concept will be intuitively clear. In addition, because the two concepts generate distinctive behavior, observed choices can also be used to determine which concept applies.[19]

As the proofs in Appendix C establish, Propositions 3 through 6 derived above for PPE also apply to CPE. Because anticipated risk exposes the decision maker to sensations of loss, she is first-order risk averse. Furthermore, expecting risk reduces her aversion to additional risk. This latter result follows partly from the force behind our analogous results above, that, when expecting and facing stochastic losses to start with, taking further risk does not increase exposure to losses as much. When a person chooses both her reference and outcome lotteries, there is an additional, parallel force acting in the same direction: when expecting and facing stochastic losses to start with, the expectation of further risk does not increase exposure to losses as much.

Despite these similarities, CPE is consistent with risk aversion that is qualitatively different, not only from standard expected-utility-over-wealth models and prospect theory, but also from the predictions of UPE, PPE, and any model where the reference point is taken as given at the moment of choice. Whereas in these models people never choose stochastically dominated options, they might do so in CPE. To illustrate this possibility, suppose μ satisfies $A3'$, and consider a lottery F that yields $w+g>w$ with probability $p\geq 0$ and w with probability $1-p$. Then $U(w+F|w+F) = [p(w+g)+(1-p)w]+[p(1-p)\mu(g)+p(1-p)\mu(-g)] = w+pg[1-(1-p)\eta(\lambda-1)]$. If $\eta(\lambda-1)>1$, which is a calibrationally

[18] CPE implicitly assumes that the decision maker maximizes the expected future sensations generated once a reference point determined by the decision is formed and outcomes are resolved. As we discuss in the conclusion, a person may not appreciate how her decision will affect future sensations of gain or loss, and she may also be influenced by current anticipatory utility. Similarly to other models of reference-dependent utility, it seems to be an appropriate first approximation to ignore these issues.

[19] For instance, suppose F_1 is a small binary lottery that generates barely enough gains for $w+F_1$ to be chosen from $\{w, w+F_1\}$, and let F_2 be a different small binary lottery with the same property. Let F be a mixture of F_1 and F_2 with equal weights. Since in PPE the reference point is fixed at the moment of choice, it is easy to show that in PPE the person would choose $w+F$ from $\{w, w+F\}$. But since in CPE a person influences the reference point by her choice, and she dislikes risky reference points, in CPE she would choose w from $\{w, w+F\}$.

plausible situation, the decision maker prefers $p = 0$ over a small $p > 0$.[20] Intuitively, raising expectations of getting g makes an outcome of no gain feel more painful. To avoid such disappointments, the person would rather give up the fragile hope of making gains. In fact, if gain-loss utility is sufficiently important, reducing exposure to sensations of loss is the decision maker's central concern.

PROPOSITION 7: *Suppose $m(\cdot)$ is linear and the decision maker faces the finite choice set D containing the deterministic outcome w and no greater deterministic outcomes. For any given $\mu_0(\cdot)$ satisfying A2, there is an $\overline{\eta}$ such that if $\mu(\cdot) = \eta\mu_0(\cdot)$ with $\eta > \overline{\eta}$, the unique CPE is to choose w.*

While the tendency to choose a stochastically dominated lottery may seem counterintuitive, it is consistent both with the flavor of some discussions in the psychology literature on welfare in risky situations, and with some experimental results on risk taking. Shane Frederick and George Loewenstein (1999) discuss, for example, how a prisoner may be made worse off by a small chance of being released, because that makes the outcome of remaining in prison much more difficult to bear. In the domain being examined in this paper, Uri Gneezy, John A. List, and George Wu (2006) find situations where risky choices are valued less than their worst possible outcome. We also feel that the preference for a stochastically dominated lottery captures in extreme form the strong risk aversion consumers display when purchasing insurance for long-term modest-scale losses, choosing low deductibles on existing insurance, and selecting expensive fixed-fee contracts for services.

There are ways in which our result must be qualified, however. The preference for dominated lotteries clearly arises only when such lotteries reduce exposure to gain-loss sensations. In addition, diminishing sensitivity can substantially reduce a person's dislike of risk for modest stakes: as the gain g in the lottery above increases, the sensation of loss from comparing nothing to g increases more and more slowly, whereas consumption utility increases linearly. And as we discuss in the conclusion, people may underappreciate how coming to expect the gain increases future sensations of loss, and hence may not be averse to choosing the lottery.

The possibility of rejecting a probabilistic gain under CPE but not under PPE reflects a more general sense in which people are more risk averse under CPE than PPE. Proposition 8 establishes that under $A3'$, if the decision maker chooses the less risky one of two lotteries in PPE, she does not choose the riskier one in CPE.

PROPOSITION 8: *Suppose $m(\cdot)$ is linear, $A3'$ holds, and for different lotteries F and F' in the decision maker's choice set, F' is a mean-preserving spread of $F + k$ for some constant k. If F is a PPE, F' is not a CPE.*

The intuition derives directly from the decision maker's desire to avoid risky expectations. That F is a PPE implies that holding the reference point fixed at F, the person prefers F to F'. When her choice affects the reference lottery in addition to the outcome lottery, as in CPE, her dislike of risky reference points leads her to prefer F to F' even more.

The most important implication of Proposition 8 is that people will be more risk averse when decisions are committed to well in advance than when people are uncommitted. But the same results also say that the availability of some risky options—exactly those that the decision maker takes in PPE but would not take in CPE—decreases welfare in PPE. Intuitively, in PPE the decision maker realizes that she will take a lottery that is attractive *fixing expectations*, and because she incorporates the possibility of good outcomes into her reference point, her sense of loss from low outcomes is increased.

Both the similarities and differences between PPE and CPE are illustrated in Table 1. Comparing the second and third panels for Gamble I, for any given amount of background risk, CPE choices are more risk averse than PPE choices. Nevertheless, CPE behavior also approaches risk neutrality with even moderate amounts of background risk.

[20] It is easy to check that $\eta(\lambda - 1) > 1$ whenever observed loss aversion is at least two-to-one—whenever overall sensitivity to losses, $1 + \eta\lambda$, is at least twice as high as overall sensitivity to gains, $1 + \eta$.

V. Immodest Risk

We now illustrate some of our model's implications for attitudes toward large-scale risk, showing that for such stakes $\mu(\cdot)$ can become largely irrelevant in determining risk preferences. Combined with the previous three sections, this means that our model can reconcile highly context-dependent behavior in modest-stakes gambles with context-independent "classical" predictions for larger stakes.

Proposition 9 shows that, when diminishing sensitivity is a significant-enough feature of gain-loss utility, the expected utility from very risky outcomes is little influenced by the reference point.

PROPOSITION 9: *Suppose* $m(\cdot)$ *has full range and* $\lim_{x\to\infty}\mu'(x) = \lim_{x\to-\infty}\mu'(x) = 0$. *For any* $r, r' \in \mathbb{R}$, $r > r'$ *and* $\varepsilon > 0$, *there is a* $\delta > 0$ *such that if F is continuously distributed with density less than* δ *everywhere, then* $0 \leq U(F|r') - U(F|r) < \varepsilon$.

If F is a very risky gamble, most of its outcomes are far from both r and r'. Since sensitivity of gain-loss utility to changes approaches zero for comparisons far apart, for a typical outcome of F it makes little difference whether it is being compared to r or r'.

Beyond this analytical result, our model's predictions for large-scale risks can be illustrated by applying the parameterized version we have considered above to larger stakes. Table 1 performs this exercise for two gambles. Gamble II is a fifty-fifty gamble that yields either \$990,000 or \$1,010,000, and Gamble III is a fifty-fifty gamble that yields either \$500,000 or \$1,500,000. While the qualitative patterns of behavior are similar for the \$100-stakes Gamble I and the \$10,000-stakes Gamble II, the sensitivity of behavior to the environment is significantly smaller in the latter case. And for the \$500,000 gamble, risk attitudes are close to what they would be without gain-loss utility, being largely determined by $m(\cdot)$ independently of the environment.

The wide range of $\tilde{\rho}$'s in Table 1 replicates an observation that an increasing number of researchers have become aware of: that the wild variation in measured risk aversion inferred by use of the prevalent reference-independent model is due to a systematic misspecification of that model, and a great deal of the variation reflects reference-dependent utility. In addition to the patterns identified in previous sections, our model predicts a systematic relationship between the inferred coefficient of risk aversion and the scale at which it is measured: when measured on large-stakes data, single- and double-digit coefficients will be found; when measured on modest-stakes data, triple-digit coefficients will be found; and when measured on small-stakes data, coefficients too embarrassingly large to report will be found. Indeed, Raj Chetty (2005) shows that existing evidence on the income elasticity of labor supply comfortably bounds the coefficient of relative risk aversion from above by two. Based on hypothetical choices between large gambles on lifetime wealth, Robert B. Barsky et al. (1997) measure an average coefficient of relative risk aversion of around five. Fitting an unemployment model with consumption and search-effort choices to data on unemployment durations, Chetty (2003) finds a coefficient of relative risk aversion of around seven. Rajnish Mehra and Edward C. Prescott (1985) estimate that to explain the historical equity premium, investors must have a coefficient of relative risk aversion well in the double digits. The deductible choices of American homeowners analyzed by Sydnor (2006) and the risk-taking behavior of Paraguayan farmers analyzed by Laura Schechter (2005) imply coefficients of relative risk aversion in the triple digits. And although risk attitudes are often measured in small-stakes laboratory experiments, coefficients calculated using the same mathematics as those above are typically not reported; if they were, they would be extremely high.[21]

VI. Caveats, Discussion, and Conclusion

Several complications limit the degree to which our model moves us to a full understanding of

[21] Many papers that report small coefficients of risk aversion for small stakes do so by dint of using the same terminology to describe mathematically different measures. By variously defining wealth as monthly income or potential income over the course of a one-hour experiment rather than lifetime wealth, these papers report figures that are orders of magnitude different from measures using wealth levels similar to those in the estimates above.

reference-dependent risk preferences. A weakness that our theory shares with other models is that it takes as one of its primitives the set of decisions and risks a person is considering, as distinct from all the decisions and risks she is facing. In fact, as Barberis, Huang, and Thaler (2006) emphasize in their setting, Propositions 2 and 6 can be interpreted as saying that if people incorporated all the risks they are facing into their expectations, reference dependence would essentially not affect risk attitudes—because people would essentially be neutral to small risks. Although psychological evidence does indicate that people often "narrowly bracket"—they isolate individual decisions and risks from relevant other decisions—relatively little is known about the extent, patterns, and effects of such bracketing phenomena.[22]

Another major limitation of our model concerns welfare. Although the analysis emphasized behavioral implications of our model, a welfare interpretation was implicit throughout and explicit at times. Insofar as the hedonic effects of choice include both gain-loss sensations and consumption utility, our utility function may provide a useful welfare measure. For two specific reasons, however, we are more hesitant about our model's welfare implications than about its behavioral implications. First, the narrow bracketing discussed above may lead people to care too much prospectively about a gain or loss whose effects are likely to be eliminated afterward by an offsetting loss or gain. Second, evidence indicates that people underestimate how quickly the reference point will adjust to a choice, and hence put too much weight on gain-loss sensations when making decisions.[23] For both these reasons, the appropriate welfare measure is likely to be closer to consumption utility than we assume in this paper. This could significantly alter some of our welfare conclusions. For example, although we showed above that the availability of a lottery can (in PPE) lower welfare defined to include gain-loss disutility, the same analysis implies that this would not be possible for welfare based solely on consumption utility.

The underestimation of changes in the reference point also has behavioral implications. If a person underestimates the effect of changes in her expectations on her preferences, she may not appreciate fully how she can rid herself of sensations of loss by ridding herself of risky expectations. Because our predictions of first-order risk aversion are driven partly by a person's desire to avoid risky expectations, this underappreciation can reduce risk aversion.

Our model also glosses over a set of issues related to what outcomes a person pays attention to. In contrast to our formal model, different outcomes resulting from the same choice often differ in salience. As noted in Sydnor (2006), for instance, having to pay for repairing an uninsured house is a very salient loss, but *not* having to pay is unlikely to result in a salient sensation of gain. A person who focuses mostly on such losses presumably has an even stronger taste for insurance than our model predicts.

Finally, our model ignores a source of utility—anticipatory emotions—that seems important in many risky situations. For instance, an investor's anxiety about funding her child's education is likely to affect both her welfare and many of her financial decisions. Insofar as anticipatory feelings are about future consumption and gain-loss utilities, our qualitative results would not be affected by adding them to the model. Quantitatively, however, anticipatory emotions can affect the degree of risk aversion; Andrew Caplin and John Leahy (2001), for instance, show that the attempt to avoid anxiety about uncertain outcomes can increase a person's preference for riskless options.

[22] For papers highlighting the role of bracketing in this and other domains, see Kahneman and Dan Lovallo (1993), Shlomo Benartzi and Thaler (1995), Daniel Read, Loewenstein, and Rabin (1999), Thaler (2000), and Barberis, Huang, and Thaler (2006).

[23] For interpretation of evidence along these lines, see Kahneman's (2003) discussion of the "transition heuristic," and Loewenstein, Ted O'Donoghue, and Rabin's (2003) formulation of "projection bias."

Appendix A: Further Definitions and Results

In this appendix we present an array of concepts and results that may be of practical use in applying our model, but that are not key to any of the main points of the paper. The first result identifies a condition such that with the reference point being the status quo, the decision maker rejects all fair gambles.

PROPOSITION 10: *Suppose $m(\cdot)$ is linear and the reference point is \$0. The decision maker rejects all fair gambles if and only if* $\lim_{x\to\infty} \mu'(-x) \geq \mu'_+(0)$.

Assumption *A3* allows the decision maker's risk lovingness in losses to be much stronger than her risk aversion in gains, which may even lead her to accept unfair gambles given a reference point of \$0. Proposition 10 says that when $m(\cdot)$ is linear, a necessary and sufficient additional condition to rule out such possibilities is that sensitivity to losses is everywhere greater than sensitivity to gains. When $m(\cdot)$ is concave, this condition is of course sufficient, but not necessary.

For our results on PPE and CPE behavior, we introduce three definitions characterizing the riskiness of lotteries. Our first definition is the conventional one of second-order stochastic dominance, except that it allows comparisons of lotteries with different means.

DEFINITION 4: *A lottery F is* less risky *than the lottery F' if F' is a mean-preserving spread of $F + k$ for some constant $k \in \mathbb{R}$.*

Because it turns out to be an especially pertinent measure of riskiness in our model, we also introduce a more specific concept that is (to our knowledge) undefined and unexplored in the literature on risk preferences.

DEFINITION 5: *The* average self-distance *of a lottery F is*

$$S(F) \equiv \iint |x - y| \;\; dF(x)dF(y).$$

The average self-distance of a lottery is the average distance between two independent draws from the lottery. A lower self-distance is a necessary but not sufficient condition for one lottery to be unambiguously less risky than another.

LEMMA 1: *If F is less risky than F', then F has lower average self-distance than F'.*

Finally, we introduce a measure for how a lottery's possible gains compare to its possible losses.

DEFINITION 6: *For a lottery F, let* $F_+ = E_F[\max\{x,0\}]$ *and* $F_- = E_F[\max\{-x,0\}]$. *The* favorability *of F is defined as $\Phi(F) \equiv 1$ if $F_+ = F_- = 0$, $\Phi(F) = \infty$ if $F_+ > 0$, and $F_- = 0$ and $\Phi(F) \equiv F_+/F_-$ otherwise.*

The favorability of a lottery is the ratio of the average gain of the lottery (relative to zero) and the average loss. Using the concepts above, Proposition 11 precisely identifies the extent of the decision maker's first-order risk aversion, generalizing Arrow's theorem to reference-dependent risky choice. This extends the limit result of Proposition 3 in the text, which said that small bets that are insufficiently better than fair will be rejected.

PROPOSITION 11: *Suppose $w \in \mathbb{R}$, and F is a lottery with bounded support.*

(i) If $\Phi(F) < (1 + \mu'_-(0))/(1 + \mu'_+(0))$, then there exists a $\bar{t} > 0$ such that for any positive $t < \bar{t}$, the unique PPE in the choice set $\{w, w + t \cdot F\}$ is to choose w.

If $\Phi(F) > (1 + \mu'_-(0))/(1 + \mu'_+(0))$, then there exists a $\bar{t} > 0$ such that for any positive $t < \bar{t}$, the unique PPE in the choice set $\{w, w + t \cdot F\}$ is to choose $w + t \cdot F$.

(ii) If $2E[F] < (\mu'_-(0) - \mu'_+(0))S[F]$*, then there exists a* $\bar{t} > 0$ *such that for any positive* $t < \bar{t}$*, the unique CPE in the choice set* $\{w, w + t \cdot F\}$ *is to choose* w*. If* $2E[F] > (\mu'_-(0) - \mu'_+(0)S[F]$*, then there exists a* $\bar{t} > 0$ *such that for any positive* $t < \bar{t}$*, the unique CPE in the choice set* $\{w, w + t \cdot F\}$ *is to choose* $w + t \cdot F$*.*

Part (i) states that when applying PPE, small bets will be accepted if and only if their favorability is greater than the "coefficient of loss aversion" associated with $u(w|r)$ —which is the slope ratio at the kink in $u(w|r)$ at $w = r$. Part (ii) says that when applying CPE, a small bet will be accepted if and only if twice its expected value is greater than the product of its average self-distance and the difference in the decision maker's sensitivity to small losses and small gains.

Proposition 12 shows that when consumption utility is linear and $A3'$ holds, we can characterize a person's CPE attitude toward a lottery purely in terms the lottery's mean and average self-distance.

PROPOSITION 12: *Suppose* $m(w) = w$ *and* $\mu(\cdot)$ *meets A3′. Then,*

(i) For any lottery F,

$$U(F|F) = E[F] - \frac{1}{2}\eta(\lambda - 1)S[F].$$

(ii) For a finite choice set D, define $y(D) = \{F \in D | \forall F' \in D, \text{ either } (a)\ S[F] < S[F'], \text{ or } (b)\ S[F] = S[F'] \text{ and } E[F] \geq E[F']\}$*. For a sufficiently high* η*,* $y(D)$ *is the set of CPE.*

Part (ii) represents a lexicographic ranking of lotteries by their lowest average self-distance, and then the highest mean. There will generally be a unique lottery with minimal average self-distance, in which case that lottery is chosen when gain-loss utility is very important—even if it is stochastically dominated by other options.

The role of average self-distance in determining a person's CPE choices when $A3'$ holds is best seen with a simple but striking observation. In a personal equilibrium of any sort, every possible sensation of gain—say from comparing an outcome x to a counterfactual $y < x$ —is matched by an equally likely and equally large loss—from comparing y to x. Because the losses are more heavily felt, net gain-loss utility will be proportional to the negative of the average of these distances.

Finally, we show two properties of risky choice in our model that it shares with the standard model. Although expected risk decreases aversion to risk under either of our solution concepts, without diminishing sensitivity it never eliminates the risk aversion completely.

PROPOSITION 13: *Suppose* $m(\cdot)$ *is linear, A3′ holds, and F second-order stochastically dominates* $F' \neq F$*. Then both the unique PPE and the unique CPE from the choice set* $\{F, F'\}$ *is to choose F.*

We also note that if F first-order stochastically dominates F', then for any given reference lottery the decision maker prefers F to F'. Hence, choosing F' cannot be a UPE. While mathematically trivial, this result is of interest because it contrasts with some of our results for CPE.

PROPOSITION 14: *Suppose the decision maker faces the choice set D. If* $F \in D$ *first-order stochastically dominates* F'*, then* F' *is not a UPE.*[24]

Appendix B: Extracting $m(\cdot)$ and $\mu(\cdot)$ from Behavior

In this appendix, we provide an algorithm that identifies a person's full utility function from her behavior (as long as the consumption utility $m(\cdot)$ inferred in the first step is continuous). Once this

[24] In fact, this result relies solely on Assumption *A1*, guaranteeing that $u(w|r)$ is increasing in w, and on no other feature of μ.

utility function is identified, our theory provides a prediction for behavior in any decision over monetary risk. To determine $\mu(\cdot)$, our algorithm works whenever $m(\cdot)$ has unbounded support; minor modifications cover the opposite case as well.

As we note in Section IV, when a person makes decisions over risk with immediate resolution and delayed consequences, the natural extension of CPE predicts that she maximizes expected consumption utility. This means that we can identify $m(\cdot)$ up to an affine transformation using standard revealed-preference techniques. Once $m(\cdot)$ is identified and a normalization is chosen (e.g., setting $m(1{,}000{,}001) - m(1{,}000{,}000) = 1$), the next two parts of our procedure allow us to infer $\mu(a)$ and $\mu(-a)$ for any given $a > 0$. Carrying out these steps for all $a > 0$ identifies the entire function $\mu(\cdot)$.

For a given a, we find wealth levels w_0, w_1, w_2 such that $m(w_2) - m(w_1) = m(w_1) - m(w_0) = a$. Then, we find the wealth level w_{CE} such that in CPE the person is indifferent between a deterministic w_{CE} and a fifty-fifty gamble that pays either w_0 or w_1. This certainty equivalent will satisfy

$$\frac{1}{2}m(w_0) + \frac{1}{2}m(w_1) + \frac{1}{4}(\mu(a) + \mu(-a)) = m(w_{CE}), \tag{5}$$

so that we have identified $\mu(a) + \mu(-a)$. Intuitively, because the ex ante evaluation of risks involved in CPE pairs any possible loss with a corresponding equally weighted gain, CPE behavior allows us to infer the difference between the pain from a loss and the pleasure from the same-sized gain. But for the same reason, CPE behavior cannot identify how large the pain and pleasure are.

To identify how big they are, we can identify the ratio of $\mu(a)$ and $|\mu(-a)|$ by eliciting the probability $p > 1/2$ such that a surprise chance p of a gain of a compensates the person for a surprise chance $1 - p$ of a loss of a. Specifically, we find the probability p such that when expecting w_1, the person is indifferent between w_1 and a gamble that pays w_2 with probability p and w_0 with probability $1 - p$. To do this, we tell her that she will get w_1 with probability $1 - \varepsilon$ and the choice with probability ε, and we let $\varepsilon \to 0$. Then, p satisfies

$$pm(w_2) + (1 - p)m(w_0) + p\mu(a) + (1 - p)\mu(-a) = m(w_1), \tag{6}$$

and if A2 is satisfied, $p > 1/2$. Equation (6) can be rearranged to show that

$$\mu(-a) = \frac{p}{2p - 1}(\mu(a) + \mu(-a)) + a;$$

$$\mu(a) = -\frac{1 - p}{2p - 1}(\mu(a) + \mu(-a)) + a,$$

which identifies $\mu(-a)$ and $\mu(a)$ based solely on the person's observed choices.

Appendix C: Proofs

PROOF OF PROPOSITION 1:

Let $F_+ = E_F[\max\{x, 0\}]$ and $F_- = E_F[\max\{-x, 0\}]$. F_+ and F_- are respectively the expected gains and losses of lottery F relative to 0.

Clearly, $U(w + F|w) \geq U(w|w)$ if and only if $(1 + \eta)F_+ \geq (1 + \eta\lambda)F_-$. Now $U(H + F|G) \geq U(H|G)$ is equivalent to

$$\iiint [w + w' + \mu(w + w' - r)]\ dG(r)dH(w)dF(w') \geq \iint [w + \mu(w - r)]\ dG(r)dH(w)$$

or

$$\iiint [w' + \mu(w + w' - r) - \mu(w - \mathrm{r})] \; dG(r)dH(w)dF(w') \geq 0.$$

Notice that for any $w' \geq 0, \mu(w + w' - r) - \mu(w - r) \geq \eta w'$, and for any $w' \leq 0, \mu(w + w' - r) - \mu(w - r) \geq \eta\lambda w'$. Hence

$$\iiint [w' + \mu(w + w' - r) - \mu(w - \mathrm{r})] \; dG(r)dH(w)dF(w')$$

$$\geq \iiint [w' + \eta \max\{w', 0\} + \eta\lambda \min\{w', 0\}] \; dG(r) \; dH(w)dF(w')$$

$$= (1 + \eta)F_{+} - (1 + \eta\lambda) F_{-} \geq 0.$$

This completes the proof.

PROOF OF PROPOSITION 2:

For each part of the proposition, we prove the first inequality ($U(H + F|G) > U(H|G)$ in part (i), and $U(H + t \cdot F|G) > U(H|G)$ in part (ii)). The other inequality is analogous.

Let F' be the mean-zero lottery that satisfies $F' + v = F$ for a constant $v > 0$.

(i) We prove that for any $\varepsilon_1 > 0$, there are $A, \varepsilon > 0$ such that if $Pr_G(r \in [k - A, k + A]) < \varepsilon$ for all $k \in \mathbb{R}$ then for any $a \in \mathbb{R}$,

$$\iint (\mu(a + w - r) - \mu(a - r)) \; dF'(w)dG(r) > -\varepsilon_1. \tag{7}$$

This is sufficient because it implies that $U(H + F'|G) - U(H|G) > -\varepsilon_1$, and hence $U(H + F|G) - U(H|G) \geq U(H + F'|G) - U(H|G) + v > v - \varepsilon_1 > 0$ for ε_1 sufficiently small.

Since μ is differentiable other than at zero and is concave in gains and convex in losses, both $\lim_{x\to\infty} \mu'(x)$ and $\lim_{x\to-\infty} \mu'(x)$ exist. This implies that for any $\varepsilon_2 > 0$, there is an A such that

$$h(b) \equiv \int (\mu(b + w) - \mu(b)) \; dF'(w) > -\varepsilon_2$$

for any $|b| > A$. h is also bounded; let its bound be M. Notice that

$$\iint (\mu(a + w - r) - \mu(a - r)) \; dF'(w)dG(r) = \int h(a - r) \; dG(r).$$

Since $Pr_G[a - r \in [-A, A]] < \varepsilon$ and $h(\cdot) > -\varepsilon_2$ outside this range, the integral above is greater than $-\varepsilon M - \varepsilon_2$. Therefore, we can choose A, ε, and ε_2 such that if G satisfies the conditions of the proposition, the integral above is greater than $-\varepsilon_1$.

(ii) Whenever $G(\cdot)$ is a continuous distribution, the function

$$h(\cdot) \equiv \int u(\cdot|r) \; dG(r)$$

is differentiable everywhere. Hence, for any $w \in \mathbb{R}$,

$$\lim_{t\to\infty} \frac{\int [h(w + t \cdot w') - h(w)] \; dF'(w')}{t} = 0$$

everywhere. Hence

$$\lim_{t\to\infty}\frac{U(H+t\cdot F'|G)-U(H|G)}{t} = \int\left[\lim_{t\to\infty}\frac{\int[h(w+t\cdot w')-h(w)]\,dF'(w')}{t}\right]dH(w)=0\cdot$$

This implies that

$$\lim_{t\to\infty}\frac{U(H+t\cdot F|G)-U(H|G)}{t}>0,$$

completing the proof.

PROOF OF PROPOSITION 3:

(The proof below establishes the result for both PPE and CPE (Definition 3 in Section IV). The proof uses the concept of average self-distance defined in Appendix A.)

Let $G = F + k$. As in the proof of Proposition 1, let $G_+ = E_G[\max\{x,0\}]$ and $G_- = E_G[\max\{-x, 0\}]$. Let $\bar{k}$ be defined by $G_+/G_- = (1+\mu'_-(0))/(1+\mu'_+(0))$. Clearly, $\bar{k}>0$. Then for any $k<\bar{k}$ we have $G_+/G_- < (1+\mu'_-(0))/(1+\mu'_+(0))$. We prove that for any such G, there is a $\bar{t}$ satisfying the statement of the proposition.

We have

$$\lim_{t\to\infty}\frac{U(w+tG|w)-U(w|w)}{t}=$$

$$\frac{\int[max\{0,tw'\}+\mu(max\{0,tw'\})]\,dG(w')+\int[min\{0,tw'\}+\mu(min\{0,tw'\})]\,dG(w')}{t}$$

$$=(1+\mu'_+(0)G_+-(1+\mu'_-(0)G_-<0.$$

Hence, there is a $\bar{t}$ such that for $t<\bar{t}$, choosing w is a UPE in the choice set $\{w, w+tG\}$.

Also:

$$\lim_{t\to 0}\frac{U(w+tG|w+tG-U(w|w)}{t}=\lim_{t\to 0}\int w'\;dG(w')+\frac{\iint\mu(t(w'-r))\,dG(w')\,dG(r)}{t}$$

$$=\int w'\;dG(w')+\lim_{t\to 0}\iint\frac{\mu(|t(w'-r)|)+\mu(-|t(w'-r)|)}{2t}\;dG(w')dG(r)$$

$$=\int w'\;dG(w')+\iint\frac{1}{2}(\mu'_+(0)-\mu'_-(0))|\,w'-r|\;dG(w')dG(r)$$

$$=\int w'\;dG(w')-\frac{1}{2}\;(\mu'_-(0)-\mu'_+(0))S(G)$$

$$\leq G_+-G_--(\mu'_-(0)-\mu'_+(0))(G_++G_-)$$

$$<(1+\mu'_+(0))G_+-(1+\mu'_-(0))G_-<0,$$

where the second-to-last inequality is true because $G_+ \geq G_-$. This establishes that w is both a PPE and CPE.

PROOF OF PROPOSITION 4:

(The proof below establishes the result for both PPE and CPE (Definition 3 in Section IV).)

By the proof of Proposition 8, if w is a UPE, then $U(w|w) > U(w + F|w + F)$, so that w is the PPE. This means that if $w + F$ is a PPE, then it must be the case that w is not a UPE. This is equivalent to $U(w + F|w) > U(w|w)$. Then, by Proposition 1, the result is immediate.

To establish the result for CPE, suppose $w + F$ is a CPE. By Proposition 8, w is not a PPE. Then, by the same logic as above, it must be the case that w is not a UPE. From here, the proof is the same as above.

PROOF OF PROPOSITION 5:

(The proof below establishes the result for both PPE and CPE (Definition 3 in Section IV).)

By a trivial modification of the proof of Proposition 1, $U(w + F|w) > U(w|w)$ implies $U(G + F|G) > U(G|G)$. Hence, $U(G + F|G) \leq U(G|G)$ implies $U(w + F|w) \leq U(w|w)$. Therefore, choosing w is a UPE in the choice set $\{w, w + F\}$. Then, by the proof of Proposition 8, $U(w|w) > U(w + F|w + F)$, so w is the PPE.

We now prove the same statement for CPE. We want to prove that if $U(G|G) \geq U(G + F|G + F)$, then $U(w|w) \geq U(w + F|w + F)$. Since w just shifts both sides of the latter inequality by a constant, it is sufficient to prove for $w = 0$. We will prove that $U(F|F) \leq U(G + F|G + F) - U(G|G)$.

Notice that the two sides are equal in consumption utility (which is equal to the expectation of F). Hence, we prove the inequality for the gain-loss-utility component. We take advantage of a geometric analogy: for any distribution H, the negative of the gain-loss utility part of $U(H|H)$ is proportional to the average self-distance of H. Therefore, the statement above is equivalent to the following: when the distribution F is added to the distribution G, the increase in the average self-distance is lower than the average self-distance of F. To show this, consider any two realizations a and b of G, and any two realizations x and y of F. By the triangle inequality,

$$|(a + x) - (b + y)| \leq |a - b| + |x - y|$$

or

$$|(a + x) - (b + y)| - |a - b| \leq |x - y|,$$

completing the proof.

PROOF OF PROPOSITION 6:

(The proof below establishes the result for both PPE and CPE (Definition 3 in Section IV).)

As in the proof of Proposition 2, for each part of the proposition we prove the first part of the statement. Also as in the proof of that proposition, define F' as the mean-zero lottery such that $F' + v = F$ for a constant $v > 0$.

(i) As an obvious implication of Proposition 2, there are $A, \varepsilon > 0$ such that the unique UPE, and hence also the PPE, is to choose $G + F$.

The proof for CPE is only slightly more complicated. By the proof of Proposition 2, for any $\varepsilon_1 > 0$ there are $A, \varepsilon > 0$ such that if $Pr_G(r \in [k - A, k + A]) < \varepsilon$ for all $k \in \mathbb{R}$, then $U(G + F'|G) - U(G|G) > -\varepsilon_1$. Applying a similar argument, there are $A, \varepsilon > 0$ such that if $Pr_G(r \in [k - A, k + A]) < \varepsilon$ for all $k \in \mathbb{R}$, $U(G + F'|G + F') - U(G + F'|G) > -\varepsilon_1$. Hence, there are $A, \epsilon > 0$ such that if $Pr_G(r \in [-A, A]) < \varepsilon$, $U(G + F|G + F) - U(G|G) = U(G + F'|G + F') - U(G|G) + v > v - 2\varepsilon_1 > 0$ for a sufficiently small ε_1.

(ii) By Proposition 2, there is a $\bar{t} > 0$ such that for $t < \bar{t}$, the unique UPE, and hence also the PPE, in the choice set $\{G, G + tF\}$ is to choose $G + tF$.

We now prove for CPE. By the same argument as in the proof of Proposition 2,

$$\lim_{t\to 0} \frac{U(G + t\cdot F|G + t\cdot F') - U(G|G)}{t} = 0,$$

so that

$$\lim_{t\to 0} \frac{U(G + t\cdot F|G + t\cdot F) - U(G|G)}{t} = v > 0.$$

PROOF OF PROPOSITION 7:

For any lottery F,

$$U(F|F) = E[F] + \eta \iint (\mu_0(w' - r)\ dF(r)dF(w') \tag{8}$$

$$= E[F] + \frac{1}{2}\eta \underbrace{\iint(\mu_0(|w' - r|) + \mu_0(-|w' - r|))\ dF(r)dF(w')}_{\equiv -\, n(F)}.$$

By A2, $n(F) > 0$ for any nondeterministic lottery F. Let $x = \min_{F\in D, F\neq w} n(F)$ and $y = \max_{F\in D, F\neq w} E[F]$. If $\eta > 2(y - w)/x$, the unique CPE is to choose w.

PROOF OF PROPOSITION 8:

We prove that if F is a UPE, then $U(F|F) > U(F'|F')$, so that F' is not a CPE. If $F' = F + k$ for some k, then the result is immediate, since in that case we would otherwise have to have $k < 0$.

That F is a UPE when F' is available implies that $U(F|F) \geq U(F'|F)$. Hence, there is a constant $k' \geq 0$ such that for $F'' = F' + k'$, we have $U(F|F) = U(F''|F)$. We prove that $U(F|F) > U(F''|F'')$, which is sufficient because $U(F''|F'') \geq U(F'|F')$.

The condition that $U(F''|F) = U(F|F)$ can be written as

$$\int w\ dF''(w) + \iint \mu(w - r)\ dF(r)dF''(w) = \int w\ dF(w) + \iint \mu(w - r)\ dF(r)dF(w). \tag{9}$$

Clearly, we must have $\int w\ dF(w) < \int w\ dF''(w)$. Otherwise, because μ is strictly increasing and concave and F'' is riskier than F, we would have $\iint \mu(w - r)\ dF(r)dF''(w) < \iint \mu(w - r)\ dF(r)dF(w)$, a contradiction.

We want to prove that

$$\int w\ dF''(w) + \iint \mu(w - r)\ dF''(r)dF''(w) < \int w\ dF(w) + \iint \mu(w - r)\ dF(r)dF(w).$$

Given equation (9), this is equivalent to

$$\iint \mu(w - r)\ dF''(r)dF''(w) < \iint \mu(w - r)\ dF(r)dF''(w).$$

Now, since the mean of F'' is greater than that of F, there is a $k'' > 0$ such that $F'' - k''$ and F have the same mean. Notice that

$$\iint \mu(w - r)\ dF''(r)dF''(w) < \iint \mu(w - r)\ d(F'' - k'')(r)dF''(w),$$

which in turn is less than or equal to

$$\iint \mu(w-r)\ dF(r)dF''(w)$$

because $F'' - k''$ is a mean-preserving spread of F and μ is concave. This completes the proof.

Proof of PROPOSITION 9:

Let $M = -\mu(-(m(r) - m(r')))$. By properties $A2$ and $A3$ of $\mu(\cdot)$, $\mu(x + m(r) - m(r')) - \mu(x) \leq M$ for any $x \in \mathbb{R}$.

Since $\lim_{x\to\infty} \mu'(x) = \lim_{x\to-\infty}\mu'(x) = 0$, for any $\varepsilon_1 > 0$, there is an A such that if $|x| > A$, then $\mu'(x) < \varepsilon_1$. Furthermore, since $m(\cdot)$ has full range, for all $\varepsilon_2 > 0$ there is a $\delta > 0$ such that if the density of F is less than δ everywhere, $Pr_F[m(w) \in [m(r') - A, m(r) + A]] < \varepsilon_2$. Denote the interval $[m(r') - A, [m(r) - A]$ by B. Under these conditions,

$$\begin{aligned} U(F|r') - U(F|r) &= \int [\mu(m(w) - m(r')) - \mu(m(w) - m(r))]\ dF(w) \\ &= \int_{m(w)\in B} [\mu(m(w) - m(r')) - \mu(m(w) - m(r))]\ dF(w) + \\ &\quad \int_{m(w)\notin B} [\mu(m(w) - m(r')) - \mu(m(w) - m(r))]\ dF(w) \\ &\leq \varepsilon_2 M + \varepsilon_1(m(r) - m(r')), \end{aligned}$$

which is less than ε for appropriately chosen $\varepsilon_1, \varepsilon_2$.

PROOF OF PROPOSITION 10:

Obvious.

PROOF OF LEMMA 1:

Since constant shifts in a distribution clearly leave the average self-distance unchanged, we can assume without loss of generality that F and F' have the same mean. Then, using that the absolute-value function is convex,

$$\begin{aligned} \iint |x-y|\ dF(x)dF(y) &\leq \iint |x-y|\ dF'(x)dF(y) \\ &= \iint |x-y|\ dF(y)dF'(x) \leq \iint |x-y|\ dF'(y)dF'(x). \end{aligned}$$

Proof of PROPOSITION 11:

Obvious from the proof of Proposition 3.

Proof of PROPOSITION 12:

Part (i) follows from equation (8). Part (ii) is trivial from part (i).

PROOF OF PROPOSITION 13:

Notice that under $A3'$, the utility function $U(\cdot|F) = \int u(\cdot|r)\ dF(r)$ is weakly concave. Hence, $U(F|F) \geq U(F'|F)$, so that choosing F is a UPE.

We now prove that $U(F|F) > U(F'|F')$, both completing the proof that F is a PPE and proving that it is a CPE. Since the expected consumption utilities cancel, this inequality is equivalent to

$$\iint \mu(w-r)\ dF(w)dF(r) > \iint \mu(w-r)\ dF'(w)dF'(r),$$

or $S(F) < S(F')$. To show this, note that by the convexity of the absolute-value function,

$$\iint |w-r|\ dF(w)dF(r) \le \iint |w-r|\ dF'(w)dF(r) = \iint |w-r|\ dF(r)dF'(w).$$

Using the same reasoning and that $F' \neq F$, the above is strictly less than

$$\iint |w-r|\ dF'(r)dF'(w) = \iint |w-r|\ dF'(w)dF'(r).$$

PROOF OF PROPOSITION 14:

For any reference lottery G, the function

$$U(\cdot|G) = \int u(\cdot|r)\ dG(r)$$

is strictly increasing. Hence, for any two distributions F, F' such that F first-order stochastically dominates F', $U(F|G) > U(F'|G)$. This implies that F' cannot be a UPE when F is available.

REFERENCES

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

Barberis, Nicholas, Ming Huang, and Richard H. Thaler. 2006. "Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case of Narrow Framing." *American Economic Review*, 96(4): 1069–90.

Barberis, Nicholas, and Wei Xiong. 2006. "What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation." National Bureau of Economic Research Working Paper 12397.

Barsky, Robert B., F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro. 1997. "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *Quarterly Journal of Economics*, 112(2): 537–79.

Bell, David, E. 1985. Disappointment in Decision Making under Uncertainty. *Operations Research*, 33(1): 1–27.

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

Birnbaum, Michael H., Gregory Coffey, Barbara A. Mellers, and Robin Weiss. 1992. "Utility Measurement: Configural-Weight Theory and the Judge's Point of View." *Journal of Experimental Psychology: Human Perception and Performance*, 18(2): 331–46.

Bombardini, Matilde, and Francesco Trebbi. 2006. Risk Aversion and Expected Utility Theory: A Field Experiment with Large and Small Stakes. Unpublished.

Bosch-Domènech, Antoni, and Joaquim Silvestre. 2006. "Averting Risk in the Face of Large Losses: Bernoulli vs. Tversky and Kahneman." Universitat Pompeu Fabra Economics Working Paper 932.

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

Breiter, Hans C., Itzhak Aharon, Daniel Kahneman, Anders Dale, and Peter Shizgal. 2001. Functional Imaging of Neural Responses to Expectancy and Experience of Monetary Gains and Losses. *Neuron*, 30(2): 619–39.

Caplin, Andrew, and John Leahy. 2001. "Psychological Expected Utility Theory and Anticipatory Feelings." *Quarterly Journal of Economics*, 116(1): 55–79.

Chetty, Raj. 2003. "Consumption Commitments, Unemployment Durations, and Local Risk Aversion." National Bureau of Economic Research Working Paper 10211.

Chetty, Raj. 2005. "Labor Supply and Risk Aversion: A Calibration Theorem." Unpublished.

Frederick, Shane, and George Loewenstein. 1999. "Hedonic Adaptation." In *Well-Being: The Foundations of Hedonic Psychology*, ed. Daniel Kahneman, Edward Diener, and Norbert Schwartz, 302–29, New York: Russell Sage Foundation.

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

Gneezy, Uri, John A. List, and George Wu. 2006. "The Uncertainty Effect: When a Risky Prospect Is Valued Less Than Its Worst Possible Outcome." *Quarterly Journal of Economics*, 121(4): 1283–309.

Gomes, Francisco. 1991. "Portfolio Choice and Trading Volume with Loss-Averse Investors." *Journal of Business*, 78(2): 675–706.

Gul, Faruk. 1991. "A Theory of Disappointment Aversion." *Econometrica*, 59(3): 667–86.

Kahneman, Daniel. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review*, 93(5) 1449–75.

Kahneman, Daniel, and Dan Lovallo. 1993. "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking." *Management Science*, 39(1): 17–31.

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

Kahneman, Daniel, and Amos Tversky. 1984. "Choices, Values, and Frames." *American Psychologist*, 39(4): 341–50.

Knetsch, Jack L., and J. A. Sinden. 1984. "Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value." *Quarterly Journal of Economics*, 99(3): 507–21.

Kobberling, Veronika and Peter P. Wakker. 2005. "An Index of Loss Aversion." *Journal of Economic Theory*, 122(1): 119–31.

Kőszegi, Botond. 2005. "Utility from Anticipation and Personal Equilibrium." Unpublished.

Kőszegi, Botond, and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics*, 121(4): 1133–65.

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-Dependent Consumption Plans." Unpublished.

Larsen, Jeff T., A. Peter McGraw, Barbara A. Mellers, and John T. Cacioppo. 2004. "The Agony of Victory and Thrill of Defeat: Mixed Emotional Reactions to Disappointing Wins and Relieving Losses." *Psychological Science*, 15(5): 325–30.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. 2003. "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics*, 118(4): 1209–48.

Loomes, Graham, and Robert Sugden. 1986. "Disappointment and Dynamic Consistency in Choice under Uncertainty." *Review of Economic Studies*, 53(2): 271–82.

Medvec, Victoria Husted, Scott F. Madey, and Thomas Gilovich. 1995. "When Less Is More: Counterfactual Thinking and Satisfaction among Olympic Medalists." *Journal of Personality and Social Psychology*, 69(4): 603–10.

Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics*, 15(2): 145–61.

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

Novemsky, Nathan, and Daniel Kahneman. 2005. "The Boundaries of Loss Aversion." *Journal of Marketing Research*, 42(2): 119–28.

Odean, Terrance. 1998. "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance*, 53(5): 1775–98.

Post, Thierry, Martijn J. Van den Assem, Guido Baltussen, and Richard H. Thaler. Forthcoming. "Deal or No Deal? Decision Making Under Risk in a Large-Payoff Game Show." *American Economic Review.*

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

Read, Daniel, George Loewenstein, and Matthew Rabin. 1999. "Choice Bracketing." *Journal of Risk and Uncertainty*, 19(1–3): 171–97.

Schechter, Laura. 2005. "Risk Aversion and Expected-Utility Theory: A Calibration Exercise." Unpublished.

Shalev, Jonathan. 2000. "Loss Aversion Equilibrium." *International Journal of Game Theory*, 29(2): 269–87.

Stone, Rebecca. 2005. "Loss Aversion and Self-Control." Unpublished.

Sugden, Robert. 2003. "Reference-Dependent Subjective Expected Utility." *Journal of Economic Theory*, 111(2): 172–91.

Sydnor, Justin. 2006. "Sweating the Small Stuff: The Demand for Low Deductibles in Homeowners Insurance." Unpublished.

Thaler, Richard H. 2000. "Mental Accounting Matters." In *Choices, Values, and Frames*, ed. Daniel Kahneman and Amos Tversky, 241–68. Cambridge; New York and Melbourne: Cambridge University Press; New York: Russell Sage Foundation.

Thaler, Richard H. and Eric J. Johnson. 1990. "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice." *Management Science*, 36(6): 643–60.