Economics

Discussion Paper

Series

EDP-2107

An Effective and Simple Tool for Measuring Loss Aversion

Olivier L'Haridon, Craig S. Webb, Horst Zank

April 2021

Economics

School of Social Sciences
The University of Manchester
Manchester M13 9PL

An Effective and Simple Tool for Measuring Loss Aversion

by Olivier L'Haridon, Craig S. Webb and Horst Zank*

April 20, 2021

Abstract. In prospect theory (PT) the loss aversion index, λ, measures the size of the concave kink of the gain-loss utility function at the reference point. A truth-telling mechanism for assessing personal beliefs, the quadratic scoring rule, is extended to measure loss aversion. We control for the bias captured by decision weights in PT and quantify λ efficiently with only three quadratic scores. In an experiment, we demonstrate these features for risk and extend the tool to ambiguity. We find median values of $\lambda = 1$ at the aggregate level for both sources of uncertainty. Probability and event weighting are less pronounced but accord with earlier findings from the literature. These weights depend on the sign of the corresponding outcomes, which is implication of reference-dependent preferences. Event weighting is also observed at the individual level. After controlling for these weights, we find very few subjects who are loss averse or gain seeking.

Keywords: Ambiguity, decision weights, loss aversion, prospect theory, scoring rules, subjective beliefs.

J*ournal of Economic Literature Classification Numbers*: C78, C91, D81, D90.

*Olivier L'Haridon: Department of Economics, Université de Rennes 1, Université Rennes, CREM UMR 6211, 7, place Hoche, 35000 Rennes, France; olharido@univ-rennes1.fr

Craig S. Webb, Department of Economics, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom; E-mail: craig.webb@manchester.ac.uk

Horst Zank, Department of Economics, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom; E-mail: horst.zank@manchester.ac.uk

During the early stages of this work, Craig Webb received financial support from the Economic and Social Research Council UK (Award Number PTA-031-2005-00117).

1 Introduction

Prospect theory (PT) is one of the most popular descriptive models for choice under risk and ambiguity (Starmer 2000, Wakker 2010, Barberis 2013). The key innovations of PT build on the intuitive assumption that outcomes of risky prospects are treated as positive and negative changes from a reference point, i.e., as gains and losses (Markowitz 1952, Kahneman and Tversky 1979). PT can therefore account for loss aversion, an aspect of risk attitude whereby, on the utility scale, a loss is more aggravating than a similar size gain is gratifying. We follow the common practice to measure loss aversion with a single parameter, λ (e.g., Tversky and Kahneman 1992, Shalev 2000, Köbberling and Wakker 2005, Peters 2012) and propose a revealed preference tool for such measurements. The surprisig finding of our experiment is that $\lambda = 1$ at both, the aggregate and individual level.

Although some of the early evidence is debated (e.g., Yechiam 2019), loss aversion has been accepted as empirically robust (for a recent contribution see Bleichrodt et al. 2020) and many economic applications build on this phenomenon (Camerer 2004, Kőszegi and Rabin 2006, 2007, Wakker 2010, Barberis 2013). Many empirical estimates, usually based on methods that invoke binary choices, are centered at the famous value $\lambda = 2$. By contrast, we measure λ under PT using choice sets that consist of specifically designed lists of prospects. First, we prove how effectively our method can be used for risk and also for ambiguity. Based on our theoretical predictions, we develop and experiment to elicit λ for these sources of uncertainty that have received considerable attention in the literature (Chew and Sagi 2008, Ergin and Gul 2009, Abdellaoui et al. 2011).

Our procedure for the elicitation of λ rests on the empirically and theoretically validated mechanism of a proper scoring rule: the popular quadratic rule (QR). As our tool directly elicits PT's decision weights for probabilities or for events and our findings, although less pronounced, accord with those reported elsewhere in the literature, our intuition for the absence of loss aversion is that earlier estimates could be procedure-dependent. Our findings question if loss aversion within PT can simply be captured by a concave kink of a gain-loss utility at the reference point.

For the theoretical setting we assume PT with linear gain-loss utility (Schmidt and Zank 2009). In this model loss aversion is revealed as a concave kink in the utility function at the reference point (Köbberling and Wakker 2005). To measure the size of this kink we proceed in two steps. First, we measure PT-decision weights in an incentive compatible way using QR-scores for events that give gains and separately for events that give losses. Here we adopt the procedure in Offerman et al. (2009). Subsequently, we correct these QR-scores using similarly elicited QR-scores for mixed prospects (i.e., prospects that feature both gains and losses). This way, bias from decision weighting can be isolated and separated from the data, such that λ can be estimated. Our theoretical analysis shows that these modifications of the standard QR-method deliver the index of loss aversion efficiently with just three elicited QR-scores. We also show that, under a duality condition, which demands that decision weights for gains and losses agree and which can be verified from our data, only two of the former three scores are required to obtain λ.

To place our experimental findings in context, we note that the literature reports some variation on values for λ. These differences result from the fact that, in the absence of a unanimously agreed upon index, various measures for loss aversion have been proposed and, unsurprisingly, the corresponding λ-values vary (see, e.g., Table 5 in Abdellaoui et al. 2007). Yet, many of these estimates are significantly larger than 1. Conceptually, there is a common agreement in the literature that loss aversion is a behavioral phenomenon attributed entirely to the treatment of outcomes (Kahneman and Tversky 1979, Rabin 2000, Shalev 2000, Neilson 2002, Köbberling and Wakker 2005, Kőszegi and Rabin 2006, Peters 2012). Accordingly, most studies take a utility-elicitation approach to determine λ. Non-parametric techniques keep the probabilities of events fixed in order to obtain a bias-free elicitation of utility from which, at a subsequent stage, loss aversion can be derived (see, for instance, Schmidt and Traub (2002) for risk, and Abdellaoui et al. (2016) for ambiguity). Such methods cleanly factor out bias that is captured by decision weights using measurement tools that build on binary choice. Binary choices are also central in studies that adopt parametric specifications for utility and decision weights under PT (e.g., Booij and van den Kuilen 2009, Gurevich et al. 2009, Krčál et al. 2016, Chapman et al. 2018).

Instead of using a sequence of binary choices, we adopt a revealed preference approach where subjects are asked to identify the best alternative from a set that contains more than just two choice alternatives. In common with all scoring rules, trade-offs are built into our mechanism. Outcomes of prospects depend explicitly on the subjective beliefs of those events, such that subjects must consider how each feasible choice jointly affects outcomes and the perceived likelihood of the event leading to that outcome. Given that the QR invokes only binary prospects, this means that subjects are asked to trade off the outcome of an event against that of its complement, such that the rate of substitution for outcomes equates the ratio of the corresponding (potentially biased, subjective) probabilities for events. Different to the traditional revealed preference method using choices from budget sets (e.g., Choi et al. 2007, Andreoni and Sprenger 2012, Ahn et al. 2014, Polisson et al. 2020), the QR dispenses of the prices that determine the budget line. From an experimental point of view, this aspect of the QR is appealing as it reduces computations and, hence, it reduces the cognitive effort required to determine the optimal choice and associated biases. The QR-method still demands trade-offs between both the probability and the utility scales (but not the otherwise exogenously fixed prices), which is key in the revealed preference approach. This renders the QR different to procedures and techniques where the trade-offs involve only the outcome scale.

Although our findings suggest that loss aversion is not the decisive feature for behavior, we find some evidence supporting probability distortion and ambiguity attitude. PT captures these biases through corresponding decision weights. Our results accord with Bruhin et al. (2010), who provide evidence that utility curvature, the classical measure of risk attitude, is not a main explanation for observed behavior for choices among small scale gain (loss) prospects under risk. By expanding the domain of prospects, we complement Bruhin et al. with evidence for mixed prospects (that simultaneously offer risky gains and losses) and we extend the evidence to include ambiguity. Our findings support the hypothesis that decision weights depend on whether they are attached to gains or to losses, which is an implication of reference-dependence in PT.[2]

One potential weakness of our method is the assumption of a linear utility on each side of the reference point. Linear utility is an important feature of scoring rules and is crucial for generating incentives for a truthful report. This aspect is known as the properness property (Savage 1971, Karni 2009, Kothiyal et al. 2011, Armantier and Treich 2013). Beyond incentive compatibility, linear utility simplifies the implementation for analysts and the resulting procedures are easy to understand for subjects. All these are desirable properties that explain why

[2]As shown in Werner and Zank (2019), having a different weighting function for probabilities of gains relative to probabilities of losses is necessary for the endogenous identification of the reference point in PT under risk. Our finding of gain–loss-dependent event weighting accords with PT and reference-dependence under ambiguity.

the QR-method is widely applied in experiments on strategic and individual decision making.[3] We maintain these appealing features and exploit them in order to obtain a simple and effective tool for the measurement of loss aversion.

Specific to our experiment, where we employ small stakes, we think that the assumption of PT with a linear gain-loss utility cannot explain the absence of loss aversion. Many studies have found that for small stakes utility is approximately linear for gains and also for losses (e.g., Fehr-Duda et al. 2006, Booij and van de Kuilen 2009, Bruhin et al. 2010, Vieider et al. 2013, Abdellaoui et al 2016) and, hence, utility curvature in those separate domains is not expected to induce much bias for elicited quadratic scores. This observation is reinforced by Rabin's (2000) theoretical demonstration that expected utility (EU) cannot offer a plausible explanation for the frequently observed small scale risk aversion, a critique that extends to several non-EU models (Neilson 2001, Safra and Segal 2008, 2009).[4] By contrast, loss aversion, would induce a strong bias. Indeed, Novemsky and Kahneman (2005, p.123) observed that risk aversion in mixed prospects involving small stakes shall be attributed exclusively to loss aversion.

We remark that, even if utility curvature would result in biased QR-scores, it is well-known how these deviations can be dealt with. For instance, Winkler and Murphy (1970), Savage (1971) or Kadane and Winkler (1988) have considered how proper scoring rules can be corrected for utility curvature. More recent studies on such corrections include Armantier and Treich (2013) and Harrison, et al. (2017).[5] For PT, where in addition to utility curvature the QR-measurements deviate from subjective beliefs due to distortions of probabilities and attitude towards ambiguity, Offerman et al. (2009) showed how one can, nevertheless, correct QR-scores and obtain reliable information. None of these studies assumed reference-dependent preferences, hence, their techniques cannot be used to explore loss aversion, the remaining bias in QR-scores

[3]Applications include, Nyarko and Schotter (2002), Armantier and Treich (2013), Bellemare et al. (2008), Rey-Biel (2009), Offerman et al. (2009), Costa-Gomes et al. (2014), Schotter and Trevino (2014), Trautmann and van de Kuilen (2015), Schlag et al. (2015), Harrison et al. (2017), Yang (2020).

[4]The recent experimental study of Bleichrodt et al. (2020) supports the resolution of the Rabin paradox to EU through reference-dependence and loss aversion.

[5]As noted by Allen (1987) one can modify the scoring rule to induce linear utility by introducing an additional randomization stage; such randomization, relies on the reduction of compound lotteries assumption, and is also invoked in the mechanisms proposed by Karni (2009) and Hossain and Okui (2013). Our QR-scoring mechanism dispenses of additional randomization stages.

under PT. We demonstrate how the QR-technique in Offerman et al. can be extended such that the bias from decision weights can simply be muted when eliciting loss aversion.

In what follows Section 2 provides some preliminaries. Section 3 introduces the traditional QR and provides extensions of it that account for reference-dependence and rank-dependence of outcomes as assumed in PT. Further, based on these extended quadratic scoring rules, we derive theoretical predictions concerning probability distortions, event weighting, duality and loss aversion. In Section 4 we present details of our experimental design, the results that test our theoretical predictions, and a brief discussion. Our findings on probability and event weighting allow for an analysis of ambiguity attitude, which we explore further in Section 5. Conclusions are provided in Section 6. Appendix A consists of proofs. Appendices B and C are supplementary. Appendix B offers a brief review on how probability distortions and events weighting are incorporated in PT and how existing methods to measure loss aversion seek to avoid those biases when eliciting λ. While this discussion is not central to our contribution, it provides some background for the interested reader and clarifies how adopting our QR-method results in efficiency gains. Appendix C contains additional figures, tables and statistical analyses and the experiment's instructions.

2 Notation

Given is a set of *states*, S, with subsets called *events*. As usual, S endowed with an algebra of events (it contains S, for each event E its complement E^c relative to S is also an event, and the union of two events is an event). Events are descriptions of a possible state of affairs, such as whether the temperature in London on a particular day will exceed 20.5 degrees centigrade or, as in our experiment, whether the price of a stock in six months from the decision point is in a certain range. Such events are sometimes referred to as naturally-occuring (Baillon, et al. 2018). An agent is uncertain if an event occurs or not.

The set of (monetary) *outcomes* is $\mathbb{R}$. The objects of choice are binary *prospects* that give an outcome conditional on an event. By $x_E y$ we denote the prospect that gives outcome x if event E occurs, otherwise it gives y (that is, when the complement, event $E^c :=$ not-E, occurs). If an event has an objective probability of occurrence, we say that the event is *risky*; in that case we write $x_{p_E} y$ where $p_E := P(E) \in [0, 1]$ is the probability of event E. The example that we consider in our experiment is an urn that contains 50 balls numbered consecutively from 1–50 but otherwise equal, where one ball is randomly drawn from the urn. Each ball is equally likely to be drawn, such that the event $E =$"the number on the ball is in the range 1–25" has a 50% chance to occur or that $P(E) = 0.5$. Here, P is a *probability measure* over events (i.e., it gives probability 1 to the universal event S and 0 to the empty event $\varnothing$, and is additive: $P(A)+P(B) = P(A\cup B)+P(A\cap B)$ for all events A and B). We also use p_E if the probability, $P(E)$, of an event E is subjective.

If probabilities are not available we have *ambiguity* and say that events are *ambiguous*. The distinction of uncertainty into risk or ambiguity was reinforced by Ellsberg (1961), leading to a growing literature on ambiguity (see Gilboa and Marinacci (2013) for a review), and has roots dating back to Keynes (1921) and Knight (1921). To capture ambiguity attitude we follow the approach of Schmeidler (1989) where ambiguity attitude is captured by general event measures. Formally an *event measure*, w, assigns weight 1 to S, 0 to the empty event ($\varnothing$) and is monotone in the sense that an event that includes another event has at least as much weight as the latter, but it is not necessarily a probability measure as it may violate additivity (e.g., $w(E) + w(E^c) \neq 1$ is permitted).

Following *prospect theory* (PT; Tversky and Kahneman 1992) we assume that monetary outcomes are treated as *gains* and *losses* relative to a *reference point*, the latter being set equal to 0. In general, the preferences of a PT-agent are *reference-dependent*, meaning that attitude towards outcomes depends on whether the latter are gains or losses and, further, attitude towards probabilities of events and towards ambiguity may be affected by whether events give gains or losses.[6] Our focus is on the index loss aversion measured at the reference point (Köbberling and Wakker 2005) and our method needs to circumvent event weighting in order to extract loss aversion from observed choices. For the purposes of this study, we formally introduce these concepts by assuming that the agent evaluates a prospect $x_E y$ by *linear prospect theory* (LPT). That is, one of the following three cases apply:

(i) if $x \geq y \geq 0$, the agent uses the event measure w^+ to evaluate a prospect by

$$LPT(x_E y) = w^+(E)x + [1 - w^+(E)]y;$$

[6]This effect of reference-dependence on event weighting has also been referred to as *sign-dependence* (e.g., Luce and Fishburn 1991).

(ii) if $x > 0 > y$, the agent uses, additionally, the event measure w^- to evaluate a prospect by

$$LPT(x_E y) = w^+(E)x + w^-(E^c)\lambda y,$$

where $\lambda > 0$ is the index of loss aversion;

(iii) if $0 \geq x \geq y$, the agent uses only the event measure w^- to evaluate a prospect by

$$LPT(x_E y) = [1 - w^-(E^c)]\lambda x + w^-(E^c)\lambda y.$$

To keep the exposition simple, we always ensure that $x \geq y$. For general purposes, we note that a similar evaluation applies for the case that the outcomes in the prospect $x_E y$ are ranked in the order $y \geq x$. LPT deviates from a general PT evaluation as it uses a linear utility function for outcomes that may have a kink at the reference point. The utility in LPT is of the linear *gain-loss* utility form

$$u(x) = \begin{cases} x, & \text{if } x \geq 0; \\ \lambda x, & \text{if } x < 0. \end{cases}$$

In LPT (and in general PT) the utility is unique up to a positive scaling parameter. That is, u can be replaced by au, for $a > 0$, without affecting the preferences of the agent over prospects. Further, the event measures under LPT are uniquely determined. The index of loss aversion, λ is uniquely determined and can, for instance, be identified from the ratio $\lambda = -u(-x)/u(x)$ for some gain x. If $\lambda > 1$ we have *loss aversion* and $\lambda < 1$ is interpreted as *gain seeking*. For details and a formal derivation of LPT from preference conditions over more general prospects we refer to Schmidt and Zank (2009).

Restricted to binary prospects, LPT includes well-known special cases. For instance, if *duality* of the event measures holds (that is, $w^+(E) + w^-(E^c) = 1$ for all events E), then LPT becomes a special case of the biseparable preference model of Ghirardato and Marinacci (2001) with a linear gain-loss utility. This means that LPT also includes Yaari's (1987) dual theory as special case (for in the latter duality holds). It is well-known that, without restrictions placed on utility functions, the biseparable preferences model is very general and includes Gul's (1991) disappointment aversion theory, Schmeidler's (1989) Choquet-expected utility for ambiguity, or the α-max-min model of Ghirardato et al. (2004) as special cases. We note that, under the duality assumption, the uniqueness results for utility are altered to cardinal uniqueness, that is, u can be replaced by $au + b$ for $a > 0$ and $b \in \mathbb{R}$. This also holds for the linear gain-loss utility in LPT if we, additionally, assume that the event measure is additive and, hence, a probability measure: this is the special case of subjective expected utility (EU) with linear gain-loss utility; subjective expected value (SEV) is a further restriction in which $\lambda = 1$ holds. These relationships indicate that, except for the commitment to linear gain-loss utility, which we argue is not very restrictive for our purposes, the class of preferences that (binary) LPT accommodates is quite rich.

3 Modified Quadratic Scoring Rules

The mechanisms considered in this study extend the *quadratic scoring rule* (QR; Brier 1950, de Finetti 1962, Hurwicz 1972, Selten 1998). An individual is asked to rate or *score* a value $r \in [0,1]$ for an event E, such that the prospect

$$(b - a[1-r]^2)_E(c - ar^2) \tag{1}$$

is most attractive. In the original QR there are no further restriction except that $a > 0$ and $b, c \in \mathbb{R}$. In an experiment, these parameters can be fixed by the designer, while the value r is chosen by the participating subject. To illustrate, set $a = b = c = 1$. Then, whatever the choice of $r \in [0,1]$, the outcomes in the prospect (1) are nonnegative, and an agent with SEV-preferences will choose r such as to maximize the expected value of the prospect. That is, the agent solves the optimization problem

$$\max_{r\in[0,1]} SEV[(1 - [1-r]^2)_E(1-r^2)] = \max_{r\in[0,1]} \{p_E[1-(1-r)^2] + (1-p_E)(1-r^2)\}, \tag{2}$$

where p_E is the agent's subjective probability of event E. As the objective function in the maximization problem (2) is concave in r, we conclude that the first-order conditions are necessary and sufficient for determining an interior solution, though the optimal value for r could be at the boundary of the probability interval (i.e., at 0 or at 1). Hence, after some basic calculations, we obtain

$$r^*_{QR}(E) = p_E \tag{3}$$

as the optimal choice for r. This conclusion remains valid if the agent views E as impossible ($p_E = 0$) or as sure ($p_E = 1$). Note that the very same implication results for any fixed values $a > 0$ and $b, c \in \mathbb{R}$. Given the property of truthful revelation of the SEV-agent's subjective probability, the QR qualifies as a proper scoring rule. The result was restated as Corollary 1 in

Offerman et al. (2009, p.1465) crediting Brier (1950) for the "first incentive-compatible result provided in the literature."

Assume now the more general LPT preferences for which we modify the QR. Consider first the case that $a > 0$ and $b, c \in \mathbb{R}$ are fixed by the designer in a way that ensures the ranking of outcome as $b - a[1-r]^2 \geq c - ar^2 \geq 0$ (whence, outcomes are gains) and reconsider the optimal choice for the score r. Similar to the case of SEV-preferences, we obtain

$$r^*_{QR^+}(E) = w^+(E) \tag{4}$$

as optimal choice for r (which also holds if $w^+(E) = 0$ or $w^+(E) = 1$). The result can also be obtained from Theorem 1 in Offerman et al. (2009) given our assumption that utility is linear for gains. The subscript "QR^+" indicates that we have modified the standard QR to account for the ranking of outcomes (i.e., event E gives the best outcome) and that all outcomes are gains. As a result, QR^+ is proper in the sense that it is a mechanism which truthfully reveals the LPT-agent's weight for the event E provided that all events of the prospect in (1) give gains and E gives the best outcome.

A similar result is obtained for our second case when the designer chooses the parameters $a > 0$ and $b, c \in \mathbb{R}$ such that $0 \geq b - a[1-r]^2 \geq c - ar^2$ whereby outcomes are losses. Now we find

$$r^*_{QR^-}(E) = 1 - w^-(E^c) \tag{5}$$

as the score for the losing event E (also valid if $w^-(E^c) = 0$ or $w^-(E^c) = 1$). Again the standard QR is modified to account for the ranking of outcomes which now are losses. As for QR^+, the properness property for the modified QR^--mechanism is maintained leading to a truthful revelation of the weight $w^-(E^c) = 1 - r^*_{QR^-}(E)$.

Combining the information derived from using equation (4) and (5) one can immediately infer if duality is violated (i.e., if $r^*_{QR^+}(E) \neq r^*_{QR^-}(E)$). The appealing feature of the (modified) QR is that with only two observations one can, in principle, test for a violation of duality in an incentive compatible way. We formulate this result as a first proposition.

PROPOSITION 1 *Assume that LPT holds. For each event E let $r^*_{QR^+}(E)$ and $r^*_{QR^-}(E)$ denote the quadratic scores for gains and losses, respectively. Duality holds if and only if*

$$r^*_{QR^+}(E) = r^*_{QR^-}(E) \tag{6}$$

for all events E. $\square$

In Appendix A we prove Proposition 1 for PT with a general utility. Assuming linear gain-loss utility then immediately provides the test proposed by Proposition 1. Our test is extremely efficient compared to earlier non-parametric tests that have used chained methods to first elicit utility before deriving the event weighting for risk (Abdellaoui 2000, Bleichrodt and Pinto 2000) and it is simpler relative to methods which have, additionally, employed matching probabilities for events to derive event weighting functions for ambiguity (e.g., Abdellaoui et al. 2005, where some empirical support for duality is reported).[7]

Next we consider the case that $a > 0$ and $b, c \in \mathbb{R}$ are fixed by the designer such that the ranking of the outcomes in prospect (1) is $b - a[1-r]^2 > 0 > c - ar^2$. This results in a further modified QR mechanism for the case that E is an *essential* event (i.e., neither E nor E^c is null or universal).[8] In that case the LPT-agent is maximizing the value of a *mixed* prospect and reports a score for the event E according to

$$r^*_{QR^M}(E) = \frac{w^+(E)}{w^+(E) + \lambda w^-(E^c)}. \tag{7}$$

From this single equation we cannot separate the information about $w^+(E)$ from $w^-(E^c)$ and λ. However, if in addition to Eq. (7) we invoke Eq. (4) and duality, we can express the loss aversion parameter λ in terms of reported scores for event E. We obtain the following result.

PROPOSITION 2 *Assume that LPT holds and that duality of the event weighting functions is satisfied. For each essential event E, the loss aversion parameter is given by*

$$\lambda = \frac{r^*_{QR^+}(E)}{1 - r^*_{QR^+}(E)} \frac{1 - r^*_{QR^M}(E)}{r^*_{QR^M}(E)} \tag{8}$$

□

The result of Proposition 2 once more demonstrates the efficient way in which the QR can be used to obtain meaningful data. It is remarkable that, under the plausible assumption of LPT and duality, only two QR-scores are required to measure loss aversion. Essentially we have

[7] The scope of those studies included the measurement of utility. To that aim, typically medium sized outcome stimuli are used in a range where utility functions can be expected to reveal some curvature. Here, we bypass the measurement of utility as we use small size outcome stimuli where utility can be expected to be approximately linear, except perhaps at the reference point where loss aversion may lead to a concave kink.

[8] In the case that E is not essential, the LPT-agent maximizes either the value of a gain prospect or that of a loss prospect, which is covered by the corresponding QR^+ and QR^- mechanisms.

provided a revealed preference tool to measure loss aversion that is simple to understand for subjects, incentive compatible, and it is implementable even in the presence of biases caused by event weighting.

For completeness we also spell out the equation for λ when duality is violated. In this case the experimenter needs to implement three QR-measurements in order to also control for deviations from duality. Combining Eqs. (4), (5) and (7) we have the following result.

PROPOSITION 3 *Assume that LPT holds. For each essential event E, the loss aversion parameter is given by*

$$\lambda = \frac{r^*_{QR^+}(E)}{1 - r^*_{QR^-}(E)} \frac{1 - r^*_{QR^M}(E)}{r^*_{QR^M}(E)}. \tag{9}$$

□

The predictions of the preceding propositions provide a basis for testable hypotheses. Hence, we can empirically demonstrate the advantages of the QR method for testing duality and for measuring loss aversion under PT by presenting an experimental study.

4 Experimental Tests of Duality and Loss Aversion

In an experiment subjects were asked to report their scores for events of binary prospects. We implemented both sources of uncertainty (ambiguity and risk) and considered all three domains of prospects (gains only, mixed, losses only). The design of the prospects was such as to employ the modified QR-mechanism in order to verify the predictions of the Propositions 1–3 and to test hypotheses that follow from those results. Although for each single implementation of a QR-elicitation we follow closely the design of Offerman et al. (2009), a comprehensive overview of our design is presented in the main text.

4.1 Design, Procedure and Stimuli

The experiment was run on computers in four sessions, each with 20–23 subjects. Ambiguous events were described as the value of a stock being in a particular range and risky events as draws of a numbered ball from an urn with the number being in a particular range. Each subject faced six ambiguous events for each of two stocks and five risky events, and was asked to report scores for $r^*_{QR^+}, r^*_{QR^M}$ and $r^*_{QR^-}$. This gave 51 questions for each individual, of which 9 questions (the three scores for each a risky and two ambiguous events) were repeated.

The block of scores for ambiguous events preceded the block of scores for risky events. Within each block, the first scores were for $r^*_{QR^+}$ followed by scores for $r^*_{QR^M}$ and finally $r^*_{QR^-}$. Within each source-mechanism-combination the order of the events were randomized. We identified four stocks from companies listed in the CAC40 of the Paris Stock Exchange and used preselected pairs of stocks (for subjects labeled as Stock 1 and Stock 2) for each of the four experimental sessions such that each group faced a different pair of stocks (to avoid information on events from one session being passed on to subjects participating in subsequent sessions). For each ambiguous event the scores of the two stocks were elicited in the order Stock 1 then Stock 2 before proceeding to scores for a new event. Within a session, all subjects faced the same pair of stocks and the order of the two stocks was fixed in advance according to the seat taken by the subject, such that any two adjacent subjects had stock specific data presented in opposite order to their neighbor.

4.2 Participants

Eighty-six subjects participated in the experiment. They were recruited via ORSEE (Greiner 2004, http://www.orsee.org/web/) from the student body of the University of Rennes. Subjects received an e-mail message advising them to register for a specific experimental session and that they need to physically be present in the experimental laboratory at the University of Rennes. Experimental sessions took place on four occasions on September 29 and 30 (two sessions) and on October 2, 2014.

The proportion of female participants was 35%. The mean age was 19.63 years, with a standard deviation of 2. The experiment started with instructions read aloud in the room (a translation is provided in Appendix C) and each QR-score was derived through a judgements task immediately succeded by the corresponding choice task. On average it took subjects 16 minutes to complete all tasks in the experiment (12 minutes to perform the judgement tasks and 4 minutes to perform the choice tasks). The duration of the entire experiment, which included a short period with some practice questions (4 for ambiguity and 2 for risk), was not recorded. Reading out instructions and practice questions took about 15 minutes and the payment after the experiment about 5 minutes.

4.3 Stimuli

This subsection presents the design of the events, the outcomes and prospects employed in the three QR-mechanisms, further descriptions of how a QR-scoring task was implemented, and the payment methods used in the experiment.

4.3.1 Ambiguous Events

The CAC40 includes the 40 largest French companies on the Paris stock exchange. The four stocks selected for the experiment were coded for data analysis purposes as Stock A (Total), Stock B (Société Générale), Stock C (LVMH short for Moët Hennessy Louis Vuitton) and Stock D (Safran SA) in the CAC40 of December 31, 2004. For subjects, the information provided in a question was a diagram showing the values of a stock price between January 1 and June 30 without mentioning the year and without any units or scaling on the axes. As argued in Offerman et al's (2009), this design generates ambiguity as the name of the stock and the year to which the observations relate were not provided in advance. At the end of the experiment subjects could verify the information provided by the experimenter as being correct.

During the experiment, a subject would be asked to provide a score for the event that the stock price on December 31 of that year is in a specified range, which are ambiguous naturally occuring events (Baillon et al. 2018) constructed as follows: (i) Event E_1 is that the stock price falls below (or is equal to) the median of annual recorded values plus one fourth of the standard deviation ($median + 1/4sd$). Event E_2 was the complement of event E_1; (ii) Event E_3 is stock price between (or equal to) the $median - 1/2sd$ and the $median + 1/4sd$. Event E_4 was the complement of E_3; and (iii) Event E_5 corresponds to the stock price being below (or equal to) the $median - 1/2sd$. Event E_6 was the complement of E_5.

Subject were not provided with median or standard deviation values. Instead, the events were displayed as a colored rectangular block on the screen next to the diagram revealing the stock price development for the preceding six months. Figure C1 in Appendix C presents the six events used in the experiment visually in a diagram. In the experiment these visualized blocks were used within the corresponding QR-scoring question.

4.3.2 Eliciting QR-Ratings for Ambiguity

For each event the three QR-scores ($r^*_{QR^+}, r^*_{QR^M}$ and $r^*_{QR^-}$) were elicited using choice sets generated by the corresponding modified QR-mechanism. To avoid potential biases (such as subjects always starting at the top of a list), we implemented the two-step elicitation procedure of Offerman et al. (2009), in which an initially judged probability is adjusted in a subsequent choice task. Following the inspection of a diagram showing variation in price values of the stock in the past six months, a subject had to make an initial guess about the probability that the stock will have a value in event E_k ($k \in \{1, \ldots, 6\}$). They would enter their probability guess in %-points out of 100 into a box on the computer screen. This is the *judgement task* within a QR-scoring question.

Based on the probability guess in a judgement task, the second elicitation step in a QR-score question consists of a *choice task*. It requires an adjustment of the initial probability guess by reporting a QR-score. For this purpose, on the computer screen a table appeared with three columns. The first column contained numbers ranging from 0–100% (in 1% point units) while the second and third column contained outcomes of a QR-prospect corresponding to those numbers. The row with the number corresponding to the initially guessed probability and the corresponding outcomes was highlighted on the screen. Further, on the screen we provided a smaller window with the stock price development and event used in the corresponding judgement task. An example of a judgement task followed by the corresponding choice task is displayed in Figures C2 and C3 of Appendix C.

The outcomes in a QR-prospect (i.e., $(b - a[1 - r]^2)_E(c - ar^2)$) were generated according to the three modified QR-mechanisms as follows: (i) for $r^*_{QR^+}$, outcomes were generated by choosing the parameters $a = 20,000$, $b = 60,000$ and $c = 20,000$; (ii) For $r^*_{QR^M}$ we generated the outcomes using $a = 20,000$, $b = 20,000$ and $c = -20,000$; and (iii) for $r^*_{QR^-}$ the outcomes were determined using $a = 20,000$, $b = -20,000$ and $c = -60,000$. By varying r from 0 to 1, these stimuli generated a list of choice alternatives in which the outcomes conditional on event E_k, $k = 1, \ldots, 6$, increased with the event weight given to E_k, while the outcomes conditional on the complement of E_k decreased with the event weight of E_k. Further, the outcomes conditional on the rated event were always ranked above the outcomes in the complementary event. In the instructions, subjects were informed that outcomes were converted into real currency according to the exchange rate 1 Euro $= 5,000$ experimental outcome points.

4.3.3 Events and QR-scores for Risk

To describe the risky events we used a bingo-cage that contained balls numbered consecutively from 1–50. This device was available for inspection in the room where the experiment took place. A risky event referred to a single draw of a ball from the cage and the number of the drawn ball being in a specified range. We used five probabilities framed as the following events: the displayed range was $\{1, 2, 3\}$ for probability $p = 0.06$; $\{1, \ldots, 15\}$ for $p = 0.30$; $\{1, \ldots, 25\}$ for $p = 0.50$; $\{1, \ldots, 35\}$ for $p = 0.70$; and $\{1, \ldots, 47\}$ for probability $p = 0.94$. Except for the framing of events, the judgement and choice tasks for risk have a similar display as for ambiguity (see Appendix C, Figures C4 and C5). To compare our results on probability weighting and on loss aversion to existing results in the literature, the QR-scores for risky events are particularly useful. If the loss aversion parameter is independent of the source of uncertainty (i.e., or risk or ambiguity) we should expect to see similar results on loss aversion for risk as for ambiguity.

4.3.4 Payment and Incentives Scheme

Each subject received a show-up fee of 4 Euros. In addition to the show-up fee, subjects received an initial endowment of 16 Euros, from which they could win or lose some amount depending on their choices during the experiment and depending on one randomly selected choice task. This task was placed in a sealed envelope and selected at the beginnig of the experiment, while at the end of the experiment the content of the envelope was opened such that the choice of the selected QR-task could be played out for real. This random incentive procedure is known as the PRINCE-method (Johnson et al. 2021). For choice under ambiguity, Baillon et al. (2015) provided arguments for implementing the random incentive system in this way. Given the outcome stimuli used in the QR-scoring questions, a subject could gain or lose up to the maximum of 16 Euros. The average payment was 19.49 Euros, with a standard deviation of 7.38 (median payment was 20.90 Euros; first quartile: 14.59; and third quartile: 23.84). Further details on the payments are provided in Appendix B.

4.4 Results

This section first reports findings on choice consistency. As differences between the sources of uncertainty can be expected, the information on consistency will inform us about the frequency of potential errors. The findings on duality and loss aversion are presented at the aggregate level and, for robustness purposes, also at the individual level. For loss aversion we report findings by initially assuming the duality assumption and state if the results are robust to dropping this assumption.

4.4.1 Consistency

Each individual participant has repeated nine QR-questions of which six were for ambiguity (one for each score and each stock) and three for risk (one for each score). We computed averages of the difference between the original QR-score and the repeated QR-score across all repeated questions, and we grouped these according to the source of uncertainty (ambiguity or risk) and according to the domain of prospects (gain, mixed, and loss). Column 2 of Table 1 reports these averages and it supplements them with percentages of consistent answers to a scoring task according to three categories: no difference between scores (Column 3), 5%-point difference (Column 4) and 10%-point difference (Column 5).

Consistency	Average difference	% no diff.	% 5 points diff.	% 10 points diff.
Overall	-0.53	36.95	52.33	68.48
Ambiguity	-1.36	19.38	37.60	58.33
Risk	1.14	72.09	81.78	88.76
$r^*_{QR^+}$	-1.47	34.88	45.35	62.02
$r^*_{QR^M}$	-0.25	37.98	55.43	69.38
$r^*_{QR^-}$	0.14	37.98	56.20	74.03

Table 1: Difference in ratings and percentage consistent answers

The results in Table 1 suggest that consistency of QR-scores was higher for risk than for ambiguity. Indeed, less than 20% of repeated QR-scores under ambiguity agreed with the original score, while for risky events more that 70% of scores were in agreement. This large difference is source-driven and is expected as the QR-questions for risk are simpler to assess and can easily be computed from the information provided, whereas the QR-questions for ambiguity were subjective assessments provided on the basis of very little information (as expected for ambiguity). If we allow for some error between initial and repeated scores, the difference in consistent scores across sources of uncertainty becomes smaller but does not disappear.

There appears to be higher consistency for QR-scores as we move from the domain of gain prospects to mixed and loss prospects. As the QR-scores for ambiguity preceded those for risk, beyond source-dependence, learning as the experiment progresses could be an alternative explanation for higher consistency for QR-scores under risk. Also, the QR-scores for gain prospects were always succeeded by the QR-scores for mixed prospects and, in turn, followed by scores for loss prospects. The somewhat higher consistency for the latter scores could, therefore, be a result of this design feature. A statistical analysis (Appendix C) delivered no significant effect of learning from the order in which the QR-scores were elicited. There was some effect attributable to the source of uncertainty (significant only at the 10%-level) with higher consistency of scores for risk than for ambiguity, as expected. Given this finding, we report subsequent results at the exact tests proposed in Propositions 1–3 and, as a robustness criterion, also with some tolerance of 5% or 10% to account for the potentially larger errors in scores under ambiguity.

4.4.2 Aggregate Raw Data on Duality

This subsection presents a summary of the aggregate data on duality. Additional tables for ambiguous events and individual stocks and for risky events are provided in Appendix C. Table 2 gives a summary of the distributions of QR-score differences for events of gains and losses across sources of uncertainty. In particular, Table 2 shows that under ambiguity very few choices satisfy duality exactly and that residual choices are distributed roughly even on either side of the exact equality test (Appendix C confirms this finding at the level of each stock).

Source/ Ratings	$r^*_{QR^+} < r^*_{QR^-}$	$r^*_{QR^+} = r^*_{QR^-}$	$r^*_{QR^+} > r^*_{QR^-}$
Overall	34.20	23.12	42.68
Ambiguity	40.50	13.08	46.41
Risk	19.07	47.21	33.72

Table 2: Order of $r^*_{QR^+}$ and $r^*_{QR^-}$ ratings in %

Under risk, nearly half of the choices satisfy the exact duality test and among the remaining choices there appears to be a slight skew towards responses where $r^*_{QR^+}$ dominates $r^*_{QR^-}$.

If we toleratee for up to 5% or 10% deviation from the exact duality condition, we obtain the distribution in Table 3. It shows that, on a scale ranging from -100% to $+100\%$, the average difference in the relevant scores is small and, in accordance with Table 2, is smaller for ambiguity than for risk due to the fraction of skewed responses for risk where $r^*_{QR^+}$ dominates $r^*_{QR^-}$. The skew in the distribution of differences in QR-scores for risk seems to be caused

Source / Ratings	Average diff.	% no diff.	% 5 points diff.	% 10 points diff.
Overall	0.74	23.12	42.34	61.90
Ambiguity	0.21	13.08	31.98	54.46
Risk	2.00	47.21	67.21	79.77

Table 3: Distribution of $r^*_{QR^+}$ and $r^*_{QR^-}$ differences with tolerance

by differences that are small, which we infer from the large fraction of scores under risk that are accommodated within a difference of up to 5% or 10% points. This can also be seen in Figure C8 in Appendix C, which provides diagrams for the corresponding distributions of these differences in scores for ambiguity and for risk.

Overall, 61.90% of score differences between $r^*_{QR^+}$ and $r^*_{QR^-}$ are equal or below 10 percentage points; for ambiguity the corresponding fraction is 54.46%, while for risk it is 79.77%. It is worth noting that, for ambiguity, these percentages are just 4–6% points lower than the corresponding percentages in Table 1 on consistency of repeated scores, which suggests that much of the observed variation could be caused by errors picked up as inconsistency in scores under ambiguity rather than genuine violations of duality. By contrast, for risk the percentages are about 9–23% points lower than those for consistency, suggesting some violation of duality at the aggregate level.

To complete the aggregate data analysis on duality, Table 4 presents results for multiple t-tests on the hypothesis that duality holds for each stock and for risk at the level of each event. The results in Table 4 show that for ambiguity there is little statistical support for a violation of duality, while for risk there is some evidence against the duality hypothesis (significant at conventional levels for $p = 0.06$ and $p = 0.50$). A similar conclusion is drawn from a Wilcoxon test (Appendix C, Table 29). We highlight the exception for risk, where we cannot reject the duality hypothesis for probability $p = 0.30$. The finding is noteworthy as the experimental literature on probability weighting reports next to no deviation from objective probabilities in the range $(0.25, 0.35)$ at the aggregate level (see, e.g., Wakker 2010). It could well be that differences in $r^*_{QR^+}$ and $r^*_{QR^-}$ scores are small because at $p = 0.30$ there is generally little bias that can be attributed to probability weighting. This finding is useful for measuring

Stock / Event	E_1	E_2	E_3	E_4	E_5	E_6
Stock A	0.501	0.151	0.790	0.670	0.584	0.809
Stock B	0.384	0.087	0.019	0.309	0.735	0.526
Stock C	0.078	0.621	0.274	0.505	0.176	0.737
Stock D	0.277	0.567	0.256	0.686	0.654	0.283
Source / Probability	0.06	0.30	0.50	0.70	0.94	
Risk	0.030	0.204	0.007	0.080	0.069	

Table 4: t-tests comparing $r^*_{QR^+}$ and $r^*_{QR^-}$

loss aversion. More generally, if one can identify events where bias attributable to probability distortions is absent, one can employ such events to elicit QR-scores for gains and mixed prospects because in LPT any difference between these scores are solely attributable to loss aversion.[9] We return to this point following our analysis on loss aversion.

4.4.3 Aggregate Data on Loss Aversion.

This part proceeds by initially assuming duality, whence a comparison of $r^*_{QR^+}(E)$ and $r^*_{QR^M}(E)$ scores for an event E is directly revealing information about loss aversion. For LPT, if $\lambda = 1$ (i.e., neither loss aversion nor gain seeking), duality implies that $r^*_{QR^+}(E) = r^*_{QR^M}(E)$ (see Proposition 2). Hence, a statistical analysis would show no significant difference between the QR-scores in the gain and mixed domains of prospects with a large portion of scores falling in the class with equality. If gain seeking ($\lambda < 1$) or loss aversion ($\lambda > 1$) holds this would imply $r^*_{QR^+}(E) < r^*_{QR^M}(E)$ or $r^*_{QR^+}(E) > r^*_{QR^M}(E)$, respectively, and similarly a large proportion of choices in those classes. Table 5 gives a summary of the distribution of the differences among the two QR-scores over the three classes: no difference, difference in the direction of loss aversion or gain seeking.

Table 5 shows that for risk half of the QR-scores are equal. For the remaining scores, there is a somewhat larger proportion of reported values that accord with loss aversion compared to gain seeking and this is independent of the source of uncertainty. Similar to Table 4, in Table 6 we

[9] In Appendix B we discuss the findings of studies that use binary choices among risky prospects with equally likely gains and losses. Such studies do not control for probability weighting and their reported measures of loss aversion are sensitive to the probability of events, reiterating the need for circumventing bias attributed to event weighting.

Source vs. Difference	$r^*_{QR^M} > r^*_{QR^+}$	$r^*_{QR^M} = r^*_{QR^+}$	$r^*_{QR^M} < r^*_{QR^+}$
Overall	34.75	25.51	39.74
Ambiguity	40.21	15.21	44.57
Risk	21.63	50.23	28.14

Table 5: Distribution of $r^*_{QR^+}$ and $r^*_{QR^M}$ differences

report results for multiple t-tests on the hypotheses that $r^*_{QR^+} = r^*_{QR^M}$ holds for each stock and for risk at the level of each event (a Wilcoxon test is provided in the Appendix C, Table 30). Table 6 supports the view that for most ambiguous events there is no statistically significant

Stock/Event	E_1	E_2	E_3	E_4	E_5	E_6
Stock A	0.027	0.169	0.914	0.728	0.500	0.180
Stock B	0.772	0.346	0.200	0.282	0.706	0.407
Stock C	0.253	0.938	0.749	0.182	0.020	0.788
Stock D	0.156	0.713	0.054	0.647	0.474	0.178
Source / Probability	0.06	0.30	0.50	0.70	0.94	
Risk	0.173	0.614	0.277	0.086	0.986	

Table 6: t-tests comparing $r^*_{QR^+}$ and $r^*_{QR^M}$

difference between QR-scores in the domain of gain and mixed prospects. This observation extends to the risky events. The exceptions are for E_1 in Stock A and E_5 for Stock C (and maybe for E_3 in Stock D). These finding would be seen as natural if differences in the scores are normally distributed around 0. Figure C9 in Appendix C supports this view.

If we drop the assumption of duality, an analogue of Table 6 can be obtained. The corresponding test compares the values $r^*_{QR^M}(1 - r^*_{QR^-})$ and $r^*_{QR^+}(1 - r^*_{QR^M})$, and essentially corresponds to a test of the (null) hypothesis that $\lambda = 1$. Except for Stock A, such a two-sided test provides p-values around 0.6, which means that there is insufficient statistical evidence to reject the null hypothesis at conventional levels.

While we expect a less spread distribution under ambiguity, as suggested by Table 5, the magnitude of the difference in the scores under ambiguity may be rather small or (in the case of violations of duality) concentrated at specific events. To account for such variation we proceed with the derivation of loss aversion parameters based on the relevant scores for each of the events.

4.4.4 Loss Aversion Parameters at the Aggregate Level

Here we provide results on the derived loss aversion parameters at the aggregate level. We report data under the assumption that duality holds and also under the assumption that duality is not satisfied. Under duality, Proposition 2 shows that the loss aversion index, here denoted λ_d with the added subscript indicating that duality is assumed, is determined according to Eq. (8). We recall the equation for an arbitrary event E here:

$$\lambda_d = \frac{r^*_{QR^+}(E)[1 - r^*_{QR^M}(E)]}{[1 - r^*_{QR^+}(E)]r^*_{QR^M}(E)}.$$

From this equation it is clear that, if subjects report one QR-score for an event at the extremes of the probability scale (i.e., 0 or 100%), we obtain minimal and maximal values for λ_d, that is, 0 or ∞. Further, if both QR-scores for an event are equal and extreme, then λ_d would not, mathematically, be well-defined. For this reason, 51 loss aversion indices could not be computed (3.49% of all potential indices). Table 7 shows that the proportion of choices in agreement with loss aversion is somewhat higher than the proportion of choices compatible with gain seeking. The last three columns in Table 7 provide median and interquartile ranges for λ_d. These values

Source/λ_d	$\lambda_d = 0$	$0 < \lambda_d < 1$	$\lambda_d = 1$	$1 < \lambda_d < \infty$	$\lambda_d = \infty$	median	Q1	Q3
Overall	4.04	30.71	22.02	39.74	2.05	1.00	0.65	1.63
Ambiguity	4.65	35.56	12.31	44.57	1.74	1.00	0.58	1.89
Risk	2.56	19.07	45.35	28.14	2.79	1.00	1.00	1.26

Table 7: Loss aversion parameters assuming duality, aggregate results. Percentage of choices for Columns 2–6. Coefficients for the last 3 columns.

suggest that, at the aggregate level, median scores are compatible with the absence of both loss aversion and gain seeking. This is particularly so for risk. Interquartile values suggest a slight tendency towards loss aversion. This tendency can also be inferred from a comparison of entries in Columns 3 and 5 of Table 7. There we observe that, out of all scores, there are 9% more values in agreement with loss aversion instead of gain seeking (independent of the source of uncertainty).

Given the assumption of duality, one can alternatively replace $r^*_{QR^+}(E)$ by $r^*_{QR^-}(E)$ and, in combination with $r^*_{QR^M}(E)$, compute the corresponding loss aversion parameter, say λ_d^-, for each event E. This is done in Appendix C (Table 31). The results show a similar distribution to Table 7. The exception relates to risk, where now interquartile values suggest a slight tendency towards gain seeking; but, on that note, we recall that duality was not supported for all events under risk (Table 4).

If we drop the duality assumption and repeat the preceding analysis based on the loss aversion parameter derived in Proposition 3, we obtain a similar summary.[10] Relative to Table

Source/λ	$\lambda = 0$	$0 < \lambda < 1$	$\lambda = 1$	$1 < \lambda < \infty$	$\lambda = \infty$	median	Q1	Q3
Overall	4.24	37.41	13.68	41.38	1.37	1	0.67	1.50
Ambiguity	4.17	41.47	4.26	46.71	1.45	1	0.60	1.71
Risk	4.42	27.67	36.28	28.60	1.16	1	0.87	1.08

Table 8: Loss aversion parameters without duality, aggregate results. Percentage of choices for columns 2-6. Coefficients for the last 3 columns.

7, after dropping duality, in Table 8 we observe fewer scores with $\lambda = 1$. Also, for ambiguity, the proportion of median parameter values in agreement with loss aversion is just marginally larger to the proportion in agreement with gain seeking. For risk the corresponding proportions are similar.

Aggregate data does, of course, not account for individual heterogeneity. The next two subsections give an overview of how the scores of individual subjects are combined in terms of duality and according to their loss aversion parameters.

4.4.5 Individual Data on Duality

A subject has been classified as satisfying duality if 60% or more of their scores were in accordance with duality. For ambiguity this means that, out of 12 pairs of scores, 8 or more have to agree with duality, while for the five risky events this means that 3 or more pairs of scores have to agree with duality. Table 9 presents the distribution of individuals according to whether they satisfy this criterion exactly (i.e., $r^*_{QR^+}(E) = r^*_{QR^-}(E)$; left panel) and whether they satisfy duality up to 10% points difference in QR-scores (i.e., $|r^*_{QR^+}(E) - r^*_{QR^-}(E)| \leq 10\%$; right panel), where we distinguish between sources of uncertainty. The results show that nearly half of the individuals satisfy exact duality under risk and that only one individual satisfies

[10] As a result of extreme QR-ratings, 48 loss aversion coefficients could not be computed (3.28% of indices).

	Risk					
	no difference			10%-difference		
Duality	Violated	Satisfied	Total	Violated	Satisfied	Total
Ambiguity						
Violated	46	39	85	8	48	56
Satisfied	0	1	1	1	29	30
Total	46	40	86	9	77	86

Table 9: Duality, Individuals; Left: equality of scores; Right: 10%-difference allowed

exact duality under ambiguity. In the right panel of Table 9 we observe that most individuals pass the more tolerant duality criterion for scores under risk. For ambiguity approximately 2/3 of the individual do not satisfy the latter duality criterion. This finding is in contrast to the results based on aggregate data and suggests that, for ambiguity, loss aversion parameters at the level of individuals assuming duality may differ to those without this assumption, while for risk the duality assumption can be justified.

4.4.6 Individual Loss Aversion Data under Duality

In this part duality is assumed although, as the analysis in the preceding subsection suggests, for ambiguity this assumption may not be justified at the level of individuals. While there are some qualitative differences when the duality assumption is dropped, the findings on loss aversion at the individual level are in many ways similar (and, hence, postponed to Appendix C). Given that each individual has rated six ambiguous events for two type of stocks and five risky events, we can compute 17 individual loss aversion parameters, one for each event. We aggregated these parameters using two simple, alternative rules. For each subject we determined the median and the mean loss aversion parameters across events for ambiguity and separately for risky events, and we supplemented these estimates with statistical analyses.

For ambiguity, the median of subjects' median λ_d-values is equal to 1, with an interquartile range (IQR) of $[0.855, 1.268]$. If we exclude values that are extreme (i.e., 0 or ∞), the median remains unchanged while the IQR is $[0.885, 1.306]$. The average of the individual medians (after excluding extreme values) is equal to 1.239. Somewhat lower values are found for λ_d^- parameters. Table 10 provides details, including Wilcoxon and t-tests for the hypothesis that

	λ_d		λ_d^-		
	all	no extr.	all	no extr.	$\lambda_d = \lambda_d^-$
Median	1.000	1.000	1.000	1.000	
Wilcoxon	0.096	0.013	0.993	0.336	0.251
Wilcoxon, log	0.147	0.029	0.788	0.467	
Q1	0.855	0.885	0.778	0.786	
Q3	1.268	1.306	1.230	1.258	
Mean	∞	1.239	∞	1.091	
t-test		0.018		0.081	0.121
t-test, log		0.030		0.473	

Table 10: Ambiguity & Duality: Median loss aversion parameters, with/without extreme values

the median and mean parameter values are equal to 1 and that $\lambda_d = \lambda_d^-$. As the distribution of parameters is restricted to positive values, we also considered the statistical tests using log-parameter values. Further analyses at the level of stock and for averages of subjects' median loss aversion parameters are provided in Appendix C (Tables 32 and 33).

Next we report findings for risky events. There, the median of subjects' median λ_d-values is also equal to 1, with an IQR of $[1, 1.142]$. If we exclude values that are extreme (i.e., 0 or ∞), the resulting statistics are virtually identical. After exclusion of extreme values, the average of the individual medians is equal to 1.101. Table 11 supports the hypothesis that $\lambda_d = \lambda_d^-$ which, in contrast to the aggregate data analysis, suggests that duality is satisfied for risk at the level of individuals.[11] Table 12, based on subjects' average of median loss aversion parameters, also suggests that many individuals do not exhibit loss aversion or gain seeking for risk when duality is assumed. That said, to some degree loss aversion is supported by the IQR's and the fact that some individuals are extremely loss averse, thereby pushing the mean of the averages above the prominent value for loss aversion of 2.

To complete the results for loss aversion parameters under the duality assumption, we first note that, at conventional levels a paired t-test shows no significant difference between risk and ambiguity in individual mean parameters (two-sided test: p-value$= 0.411$); this is confirmed

[11]It can be inferred from the proof of Proposition 1 that $\lambda_d = \lambda_d^-$ indirectly supports the view that utility curvature has no significant effect on our elicited index of loss aversion. Further arguments supporting the assumption of linear utility on each side of the reference point are presented in Appendix B.

	λ_d		λ_d^-		
	all	no extr.	all	no extr.	$\lambda_d = \lambda_d^-$
Median	1.000	1.000	1.000	1.000	
Wilcoxon	0.032	0.011	0.195	0.914	0.023
Wilcoxon, log	0.057	0.022	0.943	0.675	
Q1	1.000	1.000	1.000	1.000	
Q3	1.142	1.141	1.000	1.000	
Mean	∞	1.101	∞	1.011	
t-test		0.009		0.666	0.013
t-test, log		0.062		0.717	

Table 11: Risk & Duality: Median loss aversion parameters, with/without extreme values

	λ_d		λ_d^-		
	all	no extr.	all	no extr.	$\lambda_d = \lambda_d^-$
Median	1.050	1.018	1.000	1.000	
Wilcoxon	0.011	0.001	0.944	0.316	0.049
Wilcoxon, log	0.016	0.001	0.707	0.248	
Q1	0.915	0.924	0.849	0.919	
Q3	1.877	1.492	1.135	1.123	
Mean	∞	2.05	∞	5.282	
t-test		0.182		0.256	0.280
t-test, log		0.001		0.042	

Table 12: Risk & Duality: Mean loss aversion parameters, with/without extreme values

by a Wilcoxon test (two-sided test: p-value$= 0.893$).[12] Second, we provide a distribution of individuals as gain seeking ($r^*_{QR^+}(E) < r^*_{QR^M}(E)$), loss averse ($r^*_{QR^+}(E) > r^*_{QR^M}(E)$), neutral ($r^*_{QR^+}(E) = r^*_{QR^M}(E)$) or else unclassified, according to the criterion by which 60% or more pairs of relevant QR-scores are in the corresponding class and conditional on the source of uncertainty. We obtain Table 13, where we observe that under risk the majority of individuals

[12]There is no theoretical basis for us to assume loss aversion as sole alternative hypothesis, hence we report two-sided tests. On a different point, we remark that a paired test does not require a logarithmic transformation of the data, because there is no reason for the sampling distribution of the difference to be asymmetric.

	Risk				
	Gain Seeking	Loss Averse	Loss Neutral	Unclassified	Total
Ambiguity					
Gain Seeking	1	2	3	0	6
Loss Averse	3	4	6	1	14
Neutral	0	0	2	0	2
Unclassified	7	15	35	7	64
Total	11	21	46	8	86

Table 13: Loss aversion under duality: individual classifications

exhibit neutral loss attitude and there are nearly twice as many loss averse individuals as there are gain seeking; very few individuals cannot be classified. By contrast, under ambiguity, most individuals are unclassified and of the few classified individuals there are more than twice as many loss averse individuals as there are gain seeking.

If we allow for a 10% point difference between relevant QR-scores for gain and mixed prospects, we obtain a similar table with most individuals (78 out of 86) regarded as loss neutral for risk and just 4 as unclassified. For ambiguity the number of loss neutral subjects increases to 37 and the unclassified to 46. Very few individuals classify as loss averse or gain seeking at the 10% tolerance level (see Appendix C, Table 34). As reported in Appendix C, when the duality assumption is dropped, the corresponding results on loss aversion at the individual level are only marginally affected.

4.5 Discussion

Our approach to measure loss aversion was to elicit decision weights for risky and ambiguous events over three types of prospects (gain, mixed, and loss), and to combine these weights such that the bias attributed to decision weighting cancels out and the index of loss aversion can be identified from the collected data. We have reported that duality is not rejected at the aggregate level but that for ambiguity it may not hold at the individual level. As a result, we explore event weighting under ambiguity in more detail in the next section. Here we discuss our findings on loss aversion. Further aspects of our QR-method and on consistency are provided in Appendix B.

Our data shows little support for loss aversion at the aggregate, neither at the individual level nor across sources of uncertainty. Only if we take averages over each individual's median loss aversion parameter, we find values that give some support for loss aversion under ambiguity. Then $\lambda = 1.296$ (assuming duality, $\lambda_d = 1.484$ and $\lambda_d^- = 1.326$). For risk there is no support for loss aversion ($\lambda = 1$, $\lambda_d = 1.05$, $\lambda_d^- = 1$). As individuals who are gain seekers have parameters limited to the $(0,1)$-interval, such averages are driven mainly by those individuals with a high loss aversion parameter. Our findings clearly deviate from most existing studies (which we discuss further in Appendix B), but they agree with those of Chapman et al. (2018), von Gaudecker et al. (2011) and Gächter et al. (2010), where it is observed that the distribution of loss aversion parameters include many subjects with $\lambda < 1$, and with Brooks et al. (2014), where a model-free design using binary choices over mixed (mostly non-binary) prospects finds very few individuals who classify as loss averse. Given that, for ambiguity, the average over individual median λ_d-values are most supportive of the loss aversion hypothesis, but the corresponding λ-values are much lower, the accumulated evidence suggests that event weighting under ambiguity may partly be responsible for this difference.[13]

A specific comment concerns the practical measurement of loss aversion. In our experiment we collected many scores for various events and their complements, in order for us to have sufficient data for robustness checks. If one is only interested in a median value for loss aversion for an entire population, our data supports a more pragmatic approach. For instance, we can focus on QR-scores for an event with objective probability $p = 0.3$. Then, with only three QR-scores, one for each domain (gain, loss and mixed) of prospects, we are able to obtain a pragmatic estimate for loss aversion, say $\lambda_{p=0.3}$, based on Proposition 3. Specifically, we have

$$\lambda_{p=0.3} = \frac{r^*_{QR^+}(0.3)[1 - r^*_{QR^M}(0.3)]}{[1 - r^*_{QR^-}(0.3)]r^*_{QR^M}(0.3)}.$$

Given that for our data we observed the median scores of $r^*_{QR^+}(0.3) = r^*_{QR^-}(0.3) = 0.3 = p$, we are not surprised to find that this specific loss aversion index under risk gives a median of $\lambda_{p=0.3} = 1$ (with interquartile range $[1, 1.3]$ when all answers are included; median $\lambda_{p=0.3} = 1$ if extreme answers are excluded).

In Appendix B we discuss features of our method and experimental design further, and we also list additional findings on loss aversion from the literature. Our method finds that the

[13] This interpretation is also supported by the analysis carried out in the next section.

bias attributed to probability weighting under risk is less pronounced (see also the results of the next section). Further, by controlling for such bias we found little evidence in support of loss aversion. Given that loss aversion is regarded as one of the most "important concepts of behavioral economics" (Thaler 2016, p.1578), our finding is a concern for PT because different, theoretically equivalent, methods should not give estimates for loss aversion that are far apart. As PT was set out to be, among other features, the descriptive model for loss aversion under risk and ambiguity we would expect to see stronger support for the phenomenon. Two remarks seem relevant here: (a) our data does not reject PT, for the violations of duality are a consequence reference-dependence and they accord with PT; and (b) PT does not deliver a formal axiomatic foundation for loss aversion, and it may well be that the simple and intuitive interpretation of loss aversion as a measure for the magnitude of the kink in the utility function at the reference point has limits. Indeed, our data gives little support for this hypothesis.

5 Probabilistic Sophistication and Additive Beliefs

Probabilistic sophistication (Machina and Schmeidler 1992, Chew and Sagi 2006) holds if individuals facing choices between prospects under ambiguity assign subjective probabilities to the corresponding events. For instance, for $x \geq y$, an ambiguous event E and a probability measure P, the prospect $x_E y$ is perceived indifferent to the risky prospect $x_{p_E} y$ in which x is obtained with probability $p_E = P(E)$ and otherwise y. In the evaluation of prospects, individuals may still distort those probabilities and, hence, deviate from subjective EU. For LPT, probabilistic sophistication over gain prospects means that the weight of an event is equal that event's distorted probability. To distinguish between the holistic weighting of an event and the weighting functions for probabilities, in this section we use W^+ for the former. We write $W^+(E) = w^+(P(E))$. Abdellaoui et al. (2011) call w^+ a *source function* as it distorts subjective probabilities of ambiguous events and the deviations from subjective probabilities may, in general, depend on the source of uncertainty that generates the ambiguity.

Here, in addition to risky events, we have ambiguity resulting from just one further source of uncertainty (i.e., the price of stock on a set date being within a specific range). If we assume that objective probabilities are distorted in a similar way as subjective ones are distorted, our QR-scores for risk allow for an estimation of the weighting functions for probabilities and, when combined with the QR-scores for ambiguous events, one can identify the subjective probability of the latter (see also Offerman et al. 2009 and Kothiyal et al. 2011). Thus, our elicited QR-scores for gain prospects allow for a test of probabilistic sophistication as the following equivalences illustrate.

$$\begin{aligned} x_E y \sim x_{p_E} y \quad &\Leftrightarrow \quad W^+(E) = w^+(P(E)) \\ &\Leftrightarrow \quad r_{QR^+}(E) = w^+(P(E)). \end{aligned}$$

Assuming that the estimated probability function for gains has an inverse function, $(w^+)^{-1}$, from the latter equation we obtain

$$P(E) = (w^+)^{-1}[r_{QR^+}(E)].$$

Similarly, for prospects involving only losses, we obtain

$$P(E) = (w^-)^{-1}[1 - r_{QR^-}(E)].$$

As a result of these derivations, we can test for probabilistic sophistication by looking at the "distance from probabilistic sophistication" for the events used in our experiment, i.e., by employing the measure $I_{PS}(E) := 1 - P(E) - P(E^c)$, which we call *index of (deviation from) probabilistic sophistication*. We obtain the following straightforward result.

PROPOSITION 4 *Assume that LPT holds. Then probabilistic sophistication holds if* $I_{PS}(E) = 0$ *for all events* E*, or equivalently, if*

$$(w^+)^{-1}[r_{QR^+}(E)] + (w^+)^{-1}[r_{QR^+}(E^c)] = 1 \tag{10}$$

for all events E *attached to gains, and*

$$(w^-)^{-1}[1 - r_{QR^-}(E)] + (w^-)^{-1}[1 - r_{QR^-}(E^c)] = 1 \tag{11}$$

for all events E *attached to losses.* □

Given the preceding result, we can test for probabilistic sophistication if the probability weighting function is estimated. To illustrate, we assume the one-parameter weighting function proposed by Tversky and Kahneman (1992). It takes the form

$$w(p) = \frac{p^\gamma}{[(p^\gamma + (1-p)^\gamma]^{1/\gamma}},$$

with γ positive. We measure the index I_{PS} by combining our QR-scores for gains and for losses with the results from Proposition 4.

Based on each subject's QR-scores for five risky events for both gains and losses, the individual parameter estimate for the weighting function of gain probabilities has a median $\gamma^+ = 1$ (IQR= $[0.813, 1]$) and a mean of $\gamma^+ = 0.934$ ($sd = 0.278$). For probabilities of losses we have a median $\gamma^- = 0.972$ (IQR= $[0.820, 0.972]$) and a mean of $\gamma^- = 0.929$ ($sd = 0.199$). While there is some deviation from a linear weighting function in the direction that supports an inverse-S shape, the departure from linearity appears to be small.[14]

Table 14 summarizes the parameter-based findings on the index I_{PS} for all ambiguous events that give gains and at the level of the individual stocks. Based on parametric estimates for the

	mean	t-test	median	Wilcoxon test	Anova
Overall	0.007	0.642	0.000	0.479	0.771
Stock A	-0.016	0.605	0.000	0.890	0.693
Stock B	0.055	0.073	0.012	0.041	0.171
Stock C	-0.017	0.548	-0.020	0.308	0.622
Stock D	0.009	0.766	0.000	0.537	0.803

Table 14: Tests of probabilistic sophistication, gains, parameter based I_{PS} index

weighting function, there does not seem to be a reason to reject the assumption of probabilistic sophistication for events that give gains (the exception maybe being Stock B). For events that give losses, Table 15 indicates that the data reject the hypothesis of probabilistic sophistication at the conventional 5%-level. The consequence of this finding supports the view that event weighting is sign-dependent.

For the estimation of a one-parameter weighting function, deviations from linearity at individual QR-scores are smoothened out over all observations and some statistical error related to the parameter could propagate through to any statistical analysis of the index of probabilistic

[14] At this point a separate discussion of these findings for risk seems in order: That these estimates are larger than others found in the literature (e.g., Abdellaoui 2000) may be a consequence of asking subjects to judge objective probabilities before reporting a choice for a QR-score. If bias from utility curvature or loss aversion is small, the judgement of objective probabilities cannot be far off the actual value. As a result small deviations from objective probabilities are expected. We think it is a positive feature of this procedure to ask for an initial judgement for an event's likelihood such that it reduces biases when making a subsequent choice.

	mean	t-test	median	Wilcoxon test	Anova
Overall	-0.113	0.000	-0.066	0.000	0.000
Stock A	-0.161	0.000	-0.077	0.000	0.000
Stock B	-0.113	0.000	-0.087	0.000	0.003
Stock C	-0.079	0.002	-0.052	0.002	0.026
Stock D	-0.095	0.001	-0.021	0.000	0.011

Table 15: Tests of probabilistic sophistication, losses, parameter based I_{PS} index

sophistication I_{PS}. For this reason it seems warranted to develop robustness tests with non-parametric estimates for the probability weighting functions or, ideally, a test that directly uses the elicited QR-scores for risky events as the basis for the deviation from linearity in probabilities. The latter route is possible here as we can connect the relevant QR-scores to obtain a linearly interpolated probability weighting function. Based on the latter, we obtain a similar conclusion as for the parameter based I_{PS}-indexes (Appendix C; Tables 40 and 41).

As median data of parametric estimates suggested, probability weighting at the level of subjects did not seem very pronounced. One could, therefore, use a less complex measure to test for probabilistic sophistication in an holistic way, that is, without correcting subjective beliefs for probability weighting. In Appendix B we explores this approach and find, similar to the results in this section, support for sign-dependent event weighting.

6 Conclusion

Designing simple and efficient mechanisms to reliably measure subjective beliefs has been a central issue in many areas of economic activity (Manski 2004). Proper scoring rules are such mechanisms. One popular version, the quadratic scoring rule (QR) measures distorted beliefs under risk and ambiguity. By taking a version of prospect theory (PT) with linear gain-loss utility, we have shown how the QR can be extended to efficiently measure such beliefs when preferences display reference-dependence, and how to combine those measurements for the elicitation of the loss aversion parameter, λ, of PT. The additional benefits of the new tool are inherited from characteristic properties of scoring rules: they are relatively simple to implement, easy to understand by subjects, incentive compatible, and they dispense of chained measurements. We have demonstrated these benefits in an experiment.

While it is reassuring that our estimates for loss aversion deliver similar measures for risk as they do for ambiguity, it is surprising that the median loss aversion parameter value is $\lambda = 1$ for both sources of uncertainty. We find that preferences are reference-dependent, as assumed in PT, but that this aspect of behavior is mainly revealed through sign-dependent event weighting and to some degree also through probability weighting. For risk, the deviation from objective probabilities is less pronounced, and for ambiguity, probabilistic sophistication is violated only for losses.

PT accommodates several descriptive challenges to traditional models of uncertainty well. The model has been influential for many exciting developments of the last decades. These advances are responsible for new and more sophisticated measurement tools, yet their arrival seems to pose new challenges for PT itself. While PT delivers a framework in which reference-dependence and loss aversion can be accounted for, our data indicate that in choice situations where joint trade-offs between utility and subjective beliefs for events are explicitly invoked, loss aversion is of less relevance for revealed behavior. This finding is not necessarily a falsification of PT, but it indicates that a simple account for loss aversion as a local feature of utility (i.e., a concave kink at the reference point), while intuitive and convenient, this may not be the best way to capture the loss aversion phenomenon.

Appendix

The appendix consists of a three parts: Part A provides proofs. Parts B and C are supplementary. Part B reviews prospect theory and studies that use alternative methods to measure loss aversion, and it provides further discussion on our QR-method. Part C consists of additional material and data analyses related to our experiment.

Appendix A: Proofs

Proof of Proposition 1: For the proof we assume that PT of Tversky and Kahneman (1992) holds, with a general utility $u : \mathbb{R} \to \mathbb{R}, u(0) = 0$ strictly increasing and continuous. We further assume that utility is differentiable, except at 0 where continuity ensures the existence of left and right derivatives. Duality is defined as $w^+(E) = 1 - w^-(E^c)$ for all events E.

For the case that $E = \varnothing$ (or E is a null event that is immaterial for preferences), we have

$r^*_{QR^+}(E) = 0 = r^*_{QR^-}(E)$ and duality is verified. Similarly, for the case that $E = S$ (or E is a universal event, i.e., its complement is null), we have $r^*_{QR^+}(E) = 1 = r^*_{QR^-}(E)$ and duality is verified. Next we restrict attention to events that are essential, i.e., they are neither null nor universal.

Let $RS^+(E)$ and $RS^-(E)$ denote the ratios of marginal utility at an optimal score for an event E. That is,

$$RS^+(E) := \frac{u'[1-(1-r^*_{QR^+}(E))^2]}{u'[1-r^*_{QR^+}(E)^2]}, RS^-(E) := \frac{u'[1-(1-r^*_{QR^-}(E))^2]}{u'[1-r^*_{QR^-}(E)^2]}.$$

Following Corollary 7 in Kothiyal et al. (2011), we have

$$w^+(E) = \frac{r^*_{QR^+}(E)}{r^*_{QR^+}(E) + (1-r^*_{QR^+}(E))RS^+(E)}$$

and, similarly, one can derive

$$1 - w^-(E^c) = \frac{r^*_{QR^-}(E)}{r^*_{QR^-}(E) + (1-r^*_{QR^-}(E))RS^-(E)}.$$

Duality says that, wherever well-defined, the ratio between the latter two equations is equal to 1. Therefore, we obtain

$$\frac{r^*_{QR^+}(E) + (1-r^*_{QR^+}(E))RS^+(E)}{r^*_{QR^+}(E)} = \frac{r^*_{QR^-}(E) + (1-r^*_{QR^-}(E))RS^-(E)}{r^*_{QR^-}(E)},$$

which, after some manipulation, gives

$$\frac{RS^+(E)}{RS^-(E)} = \frac{[1-r^*_{QR^-}(E)]r^*_{QR^+}(E)}{r^*_{QR^-}(E)[1-r^*_{QR^+}(E)]},$$

for all essential events E.

Assuming LPT, means that $RS^+(E) = RS^-(E) = 1$, hence we obtain

$$\frac{1-r^*_{QR^+}(E)}{r^*_{QR^+}(E)} = \frac{1-r^*_{QR^-}(E)}{r^*_{QR^-}(E)}$$

or, equivalently, $r^*_{QR^+}(E) = r^*_{QR^-}(E)$ for all essential events E. The converse, that $r^*_{QR^+}(E) = r^*_{QR^-}(E)$ implies duality, follows immediate from the arguments in main text. This completes the proof of Proposition 1. □

Proof of Proposition 2: The proof follows from Proposition 1 when combined with Equation (7). □

Proof of Proposition 3: The proof follows from combining and simplifying the Equations (4), (5), and (7). □

The proof of Proposition 4 follows from the derivations in the main text of Section 5.

References

Abdellaoui, Mohammed (2000) "Parameter-Free Elicitation of Utility and Probability Weighting Functions," Management Science 46, 1497–1512.

Abdellaoui, Mohammed, Aurélien Baillon, Laetitia Placido, & Peter P. Wakker (2011), "The Rich Domain of Uncertainty: Source Functions and their Experimental Implementation," American Economic Review 101, 699–727.

Abdellaoui, Mohammed, Han Bleichrodt, Olivier l'Haridon, & Dennie van Dolder (2016) "Measuring Loss Aversion under Ambiguity: A Method to Make Prospect Theory Completely Observable," Journal of Risk and Uncertainty 52, 1–20.

Abdellaoui, Mohammed, Han Bleichrodt, & Corina Paraschiv (2007) "Loss Aversion under Prospect Theory: A Parameter-Free Measurement," Management Science 53, 1659–1674.

Abdellaoui, Mohammed, Frank Vossmann, & Martin Weber (2005) "Choice-Based Elicitation and Decomposition of Decision Weights for Gains and Losses under Uncertainty," Management Science 51, 1384–1399.

Ahn, David S., Syngjoo Choi, Douglas Gale, & Shachar Kariv (2014) "Estimating Ambiguity Aversion in a Portfolio Choice Experiment," Quantitative Economics 5, 195–223.

Allen, Franklin (1987) "Discovering Personal Probabilities When Utility Functions Are Unknown," Management Science 33, 542–544.

Andreoni, James & Charles Sprenger (2012) "Estimating Time Preference from Convex Budgets," American Economic Review 102, 3333–3356.

Armantier, Olivier & Nicolas Treich (2013) "Proper Scoring Rules: Incentives, Stakes and Hedging," European Economic Review 62, 17–40.

Baillon, Aurélien, Yoram Halevy, & Chen Li (2015) "Experimental Elicitation of Ambiguity Attitude Using the Random Incentive System," Mimeo, Erasmus University Rotterdam, the Netherlands.

Baillon, Aurélien, Zhenxing Huang, Asli Selim, & Peter P. Wakker (2018) "Measuring Ambiguity Attitudes for All (Natural) Events," Econometrica, 86, 1839–1858.

Barberis, Nicholas (2013) "Thirty Years of Prospect Theory in Economics: A Review and Assessment," Journal of Economic Perspectives 27, 173–195.

Bellemare, Charles, Sabine Kröger, & Arthur van Soest (2008) "Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities," Econometrica 76(4), 815–839.

Bernheim, B. Douglas & Charles Sprenger (2020) "On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting," Econometrica 88, 1363–1409.

Bleichrodt, Han, Jason N. Doctor, Yu Gao, Chen Li, Daniella Meeker, & Peter P. Wakker (2020) "Resolving Rabin's Paradox," Journal of Risk and Uncertaity 59, 239–260.

Bleichrodt, Han & José Luis Pinto (2000) "A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis," Management Science 46, 1485–1496.

Booij, Adam S. & Gijs van de Kuilen (2009) "A Parameter-free Analysis of the Utility of Money for the General Population under Prospect Theory," Journal of Economic Psychology 30, 651–666.

Brier, Glenn W. (1950) "Verification of Forecasts Expressed in Terms of Probability," Monthly Weather Review 78, 1–3.

Brooks, Peter, Simon Peters, & Horst Zank (2014) "Risk Behavior for Gain, Loss, and Mixed Prospects," Theory and Decision 77, 153–182.

Bruhin, Adrian, Helga Fehr-Duda, & Thomas Epper (2010) "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion," Econometrica 78, 1372–1412.

Camerer, Colin F. (2004) "Prospect Theory in the Wild: Evidence from the Field," Camerer, C. F., et al., (Eds.), *Advances in Behavioral Economics*. Princeton University Press, Princeton, 148–161.

Chapman Jonathan, Erik Snowberg, Stephanie Wang & Colin F. Camerer (2018) "Loss Attitudes in the U.S. Population: Evidence from Dynamically Optimized Sequential Experimentation (DOSE)," NBER Working Paper No. 25072.

Chew, Soo Hong & Jacob S. Sagi (2008) "Small Worlds: Modeling Attitudes Toward Sources of Uncertainty," Journal of Economic Theory 139, 1–24.

Choi, Syngjoo, Raymond Fishman, Douglas Gale, & Shachar Kariv (2007) "Consistency and Heterogeneity of Individual Behavior under Uncertainty," American Economic Review 97, 1921–1938.

Costa-Gomes, Miguel A., Steffen Huck, & Georg Weizsäcker (2014) "Beliefs and Actions in the Trust Game: Creating Instrumental Variables to Estimate the Causal Effect," Games and Economic Behavior 88, 298–309.

de Finetti, Bruno (1962) "Does It Make Sense to Speak of "Good Probability Appraisers"?". In Isidore J.Good (ed.) The Scientist Speculates: An Anthology of Partly-Baked Ideas, William Heinemann Ltd., London. Reprinted as Ch. 3 in Bruno de Finetti (1972) "Probability, Induction and Statistics." Wiley, New York.

Ellsberg, Daniel (1961) "Risk, Ambiguity and the Savage Axioms," Quarterly Journal of Economics 75, 643–669.

Ergin, Haluk & Faruk Gul (2009) "A Theory of Subjective Compound Lotteries," Journal of Economic Theory 144, 899–929.

Fehr-Duda, Helga, Manuele De Gennaro, & Renate Schubert (2006) "Gender, Financial Risks, and Probability Weights," Theory and Decision 60, 283–313.

Gächter, Simon, Eric J. Johnson, and Andreas Herrmann (2010) "Individual-Level Loss Aversion in Riskless and Risky Choices" CeDEx Discussion Paper Series, No. 2010-20.

Ghirardato, Paolo, Fabio Maccheroni, & Massimo Marinacci (2004) "Differentiating Ambiguity and Ambiguity Attitude," Journal of Economic Theory 118, 133–173.

Ghirardato, Paolo & Massimo Marinacci (2001) "Risk, Ambiguity, and the Separation of Utility and Beliefs," Mathematics of Operations Research 26, 864–890.

Gilboa, Itzhak, & Massimo Marinacci (2013) "Ambiguity and the Bayesian paradigm." D. Acemoglu, M. Arellano, and E. Dekel edition: in Advances in Economics and Econometrics, Tenth World Congress of the Econometric Society, Volume I, Economic Theory. Cambridge University Press.

Greiner, Ben (2004) "The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics." Working Paper, University of Cologne.

Gul, Faruk (1991) "A Theory of Disappointment Aversion," Econometrica 59, 667–686.

Gurevich, Gregory, Doron Kliger & Ori Levy (2009) "Decision-making under Uncertainty – A Field Study of Cumulative Prospect Theory," Journal of Banking and Finance 33, 1221–1229.

Harrison, Glenn W., Jimmy Martínez-Correa, J. Todd Swarthout, & Eric R. Ulm (2017) "Scoring Rules for Subjective Probability Distributions," Journal of Economic Behavior and Organization 134, 430–448.

Hossain, Tanjim, & Ryo Okui (2013) "The Binarized Scoring Rule," Review of Economic Studies 80, 984–1001.

Hurwicz, Leonid (1972) "On Informationally Decentralized Systems." In Charles Bartlett McGuire & Roy Radner (eds.) Decision and Organization, 297–336, North-Holland, Amsterdam.

Johnson, Cathleen, Aurélien Baillon, Han Bleichrodt, Zhihua Li, Dennie van Dolder, & Peter P. Wakker (2021) "Prince: An Improved Method for Measuring Incentivized Preferences," Journal of Risk and Uncertainty, forthcoming.

Kadane, Joseph B. & Robert L. Winkler (1988) "Separating Probability Elicitation from Utilities," Journal of the American Statistical Association 83, 357–363.

Kahneman, Daniel & Amos Tversky (1979) "Prospect Theory: An Analysis of Decision under Risk," Econometrica 47, 263–291.

Karni, Edi (2009) "A Mechanism for Eliciting Probabilities," Econometrica 77(2), 603–606.

Keynes, John M. (1921) A Treatise on Probability. London: Macmillan and Co.

Knight, Frank H. (1921) Risk, Uncertainty, and Profit. Boston: Houghton, Mifflin.

Köbberling, Veronika & Peter P. Wakker (2005) "An Index of Loss Aversion," Journal of Economic Theory 122, 119–131.

Kőszegi, Botond, & Matthew Rabin (2006) "A Model of Reference-dependent Preferences," Quarterly Journal of Economics 121, 1133–1165.

Kőszegi, Botond & Matthew Rabin (2007) "Reference-Dependent Risk Attitudes," American Economic Review 97, 1047–1073.

Kothiyal, Amit, Vitalie Spinu, & Peter P. Wakker (2011) "Comonotonic Proper Scoring Rules to Measure Ambiguity and Subjective Beliefs," Journal of Multi-Criteria Decision Analysis 17(3–4), 101–113.

Krčál Ondřej, Michal Kvasnička & Rostislav Staněk (2016) "External Validity of Prospect Theory: The Evidence from Soccer Betting," Journal of Behavioral and Experimental Economics 65, 121–127.

Luce, R. Duncan & Peter C. Fishburn (1991) "Rank- and Sign-Dependent Linear Utility Models for Finite First-Order Gambles," Journal of Risk and Uncertainty 4, 29–59.

Machina, Mark J. & David Schmeidler (1992) "A More Robust Definition of Subjective Probability," Econometrica 60, 745–780.

Manski, Charles F. (2004) "Measuring Expectations," Econometrica 72, 1329–1376.

Markowitz, Harry M. (1952) "The Utility of Wealth," Journal of Political Economy 60, 151–158.

Neilson, William S. (2001) "Calibration Results for Rank-dependent Expected Utility," Economics Bulletin 4, 1–5.

Neilson, William S. (2002) "Comparative Risk Sensitivity with Reference-Dependent Preferences," Journal of Risk and Uncertainty 24, 131–142.

Novemsky, Nathan & Daniel Kahneman (2005) "The Boundaries of Loss Aversion," Journal of Marketing Research 42, 119–128.

Nyarko, Yaw & Andrew Schotter (2002) "An Experimental Study of Belief Learning Using Elicited Beliefs," Econometrica 70, 971–1005.

Offerman, Theo, Joep Sonnemans, Gijs van de Kuilen, & Peter P. Wakker (2009) "A Truth Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes," Review of Economic Studies 76, 1461–1489.

Peters, Hans J.M. (2012) "A Preference Foundation for Constant Loss Aversion" Journal of Mathematical Economics 48, 21–25.

Polisson, Matthew, John K.-H. Quah, & Ludovic Renou (2020) "Revealed Preferences over Risk and Uncertainty," American Economic Review 110, 1782–1820.

Rabin, Matthew (2000) "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," Econometrica 68, 1281–1292.

Rey-Biel, Pedro (2009) "Equilibrium Play and Best Response to (Stated) Beliefs in Normal Form Games," Games and Economic Behavior 65(2), 572–585.

Safra, Zvi, & Uzi Segal (2008) "Calibration Results for Non Expected Utility Preferences," Econometrica 76, 1143–1166.

Safra, Zvi, & Uzi Segal (2009) "Risk Aversion in the Small and in the Large:Calibration Results for Betweenness Functionals," Journal of Risk and Uncertainty 38, 27–37.

Savage, Leonard J. (1971) "Elicitation of Personal Probabilities and Expectations," Journal of the American Statistical Association 66, 783–801.

Schlag, Karl H., James Tremewan, & Joël J. van der Weele (2015) "A Penny for your Thoughts: A Survey of Methods for Eliciting Beliefs," Experimental Economics 18, 457–490.

Schmeidler, David (1989) "Subjective Probability and Expected Utility without Additivity," Econometrica 57, 571–587.

Schmidt, Ulrich & Stefan Traub (2002) "An Experimental Test of Loss Aversion," Journal of Risk and Uncertainty 25, 233–249.

Schmidt, Ulrich & Horst Zank (2009) "A Simple Model of Cumulative Prospect Theory," Journal of Mathematical Economics 45, 308–319.

Schotter, Andrew, & Isabel Trevino (2014) "Belief Elicitation in the Laboratory," Annual Review of Economics 6, 103–128.

Selten, Reinhard (1998) "Axiomatic Characterization of the Quadratic Scoring Rule," Experimental Economics 1, 43–62.

Shalev, Jonathan (2000) "Loss Aversion Equilibrium," International Journal of Game Theory 29, 269–287.

Starmer, Chris (2000) "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk," Journal of Economic Literature 38, 332–382.

Thaler, Richard H. (2016) "Behavioral Economics: Past, Present, and Future," American Economic Review 106, 1577–1600.

Trautmann, Stefan T., & Gijs van de Kuilen (2015) "Belief Elicitation: A Horse Race Among Truth Serums," Economic Journal 125, 2116–2135.

Tversky, Amos & Daniel Kahneman (1992) "Advances in Prospect Theory: Cumulative Representation of Uncertainty," Journal of Risk and Uncertainty 5, 297–323.

Vieider, Ferdinand M., Lubomír Cingl, Peter Martinsson, & Hrvoje Stojic (2013) "Separating Attitudes towards Money from Attitudes towards Probabilities: Stake Effects and Ambiguity as a Test for Prospect Theory," WZB Discussion Paper No. SP II 2013-401, Berlin.

von Gaudecker, Hans-Martin, Arthur van Soest, & Erik Wengström (2011) "Heterogeneity in Risky Choice Behavior in a Broad Population," American Economic Review 101, 664–694.

Wakker, Peter P. (2010) Prospect Theory: For Risk and Ambiguity. Cambridge University Press, Cambridge, UK.

Werner, Katarzyna M. & Horst Zank (2019) "A Revealed Reference Point for Prospect Theory," *Economic Theory* 67, 731–773.

Winkler, Robert L. & Allan H. Murphy (1970) "Nonlinear Utility and the Probability Score," Journal of Applied Meteorology 9, 143–148.

Yaari, Menahem E. (1987) "The Dual Theory of Choice under Risk," Econometrica 55, 95–115.

Yang, Jingni (2020) "The Uniqueness of Local Proper Scoring Rules: The Logarithmic Family," Theory and Decision 88, 315–322.

Yechiam, Eldad (2019) "Acceptable Losses: The Debatable Origins of Loss Aversion," Psychological Research 83, 1327–1339.