EURO J Decis Process
https://doi.org/10.1007/s40070-018-0081-8

ORIGINAL ARTICLE

On proper scoring rules and cumulative prospect theory

Arthur Carvalho[1] · Stanko Dimitrov[2] · Kate Larson[3]

Received: 13 August 2017 / Accepted: 10 April 2018

Abstract Scoring rules are traditional techniques to measure the association between a reported belief and an observed outcome. The condition that a scoring rule is proper means that an agent maximizes his expected score when he reports a belief that equals his true belief. The implicit assumption that the agent is risk neutral is, however, often unrealistic, at least when the underlying agent is a human. Modern decision theories based on rank-dependent utilities, such as *cumulative prospect theory*, have been shown to be more effective at describing how human beings make decisions under risk and uncertainty. Traditional proper scoring rules are, however, incompatible with cumulative prospect theory because they fail to satisfy a property called comonotonicity. In this paper, we provide novel insights on why comonotonicity is crucial to make proper scoring rules indeed proper when eliciting beliefs from cumulative prospect theory agents. After suggesting strategies to create comonotonic proper scoring rules, we propose calibration procedures to obtain an agent's true belief by removing the influence of the agent's value function and weighting functions from his reported belief, when beliefs are elicited by means of comonotonic proper scoring rules.

✉ Arthur Carvalho
arthur.carvalho@miamioh.edu

Stanko Dimitrov
sdimitrov@uwaterloo.ca

Kate Larson
kate.larson@uwaterloo.ca

1 Department of Information Systems and Analytics, Farmer School of Business, Miami University, Oxford, OH 45056, USA

2 Department of Management Sciences, University of Waterloo, Waterloo, ON N2L 3G1, Canada

3 David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada

Published online: 25 April 2018

Keywords Proper scoring rules · Cumulative prospect theory · Comonotonicity

Mathematics Subject Classification 62C99 (Decision theory - None of the above, but in this section)

1 Introduction

There are many scenarios in which an agent's assessment of the likelihood of a future event is of interest to others. For example, a manager might be interested in the probability employees assign to the success of a new product/service. Likewise, central banks often rely on experts' forecasts of economic indicators, such as GDP and unemployment rate, when shaping economic policies. A potential issue in those scenarios is that strategic agents are not necessarily honest when reporting their beliefs. For example, Friedman (1983) said:

> In the absence of a well-chosen incentive structure, the experts may indulge in game playing which distorts their stated probability distributions. For instance, casual observation of economic forecasters suggests that experts who feel they have a reputation to protect will tend to produce a forecast near the consensus, and experts who feel they have a reputation to build will tend to overstate the probabilities of events they feel are understated in consensus." (Friedman 1983 [p. 447])

This is precisely what Nakazono (2013) found when analyzing forecasts from board members of the Federal Open Market Committee. In particular, Nakazono (2013) suggested that the reported forecasts are heavily dependent on the previous consensus: members with permanent voting rights tend to report forecasts close to the previous consensus, whereas members who rotate voting rights tend to report forecasts far away from the previous consensus. In other words, there is a mix of herding and anti-herding behavior, which implies that individual members behave strategically.

When agents report subjective probabilities (beliefs), strategic behavior is very often undesirable since the reported information might not correspond to an agent's true belief. *Proper scoring rules* (Winkler and Murphy 1968) are well-established scoring methods to induce honest reporting of subjective probabilities. Formally, an agent maximizes the expected score from a proper scoring rule by honestly reporting his belief. Hence, an implicit assumption behind the use of proper scoring rules is that agents are *risk neutral*, i.e., that they behave so as to maximize their expected scores. We discuss later in this paper how reporting a belief under a proper scoring rule is equivalent to making a decision under uncertainty. Since the paradoxes noted by Allais (1953), violations of risk neutrality and, more generally, violations of the expected utility theory framework have been widely reported in the literature (Starmer 2000). Hence, a logical research direction is to adapt proper scoring rules to a more appropriate decision theory.

Modern theories of decision under uncertainty based on rank-dependent utilities assert that the values that human beings derive from payoffs are represented by nonlinear *utility functions*, which in turn are weighted by *decision weights*, instead of subjective probabilities (Quiggin 1982; Schmeidler 1989). Decision weights are defined in terms of differences between *weighting functions* applied to cumulative probabilities. Thus, according to rank-dependent models, both a utility function and weighting functions drive an agent's attitude towards risk and uncertainty.

Cumulative prospect theory (Tversky and Kahneman 1992) is regarded as one of the most prominent theories of individual decisions under risk and uncertainty. As we elaborate in Sect. 4, cumulative prospect theory advances original rank-dependent models by considering *loss aversion* and *reference dependence*. Camerer (2004) documented the superior predictive performance of cumulative prospect theory over expected utility theory, which subsumes risk neutrality, for a range of phenomena, e.g., the equity premium puzzle, the disposition effect, asymmetric price elasticities, etc. Consider, for example, the disposition effect, i.e., the tendency of investors to sell stocks that have gone up in value since purchase as opposed to the ones that have gone down. Prospect theory explains the disposition effect by assuming that investors' preferences are defined over realized gains and losses (Barberis and Xiong 2009). That said, an investor holding a stock that has risen in value since purchase may think of the same as trading at a gain. Hence, the investor might be inclined to sell the stock if he is risk-averse over gains. Alternatively, the investor might be inclined to hold on to a stock that has gone down in value if he is risk seeking over losses. Therefore, the action taken by the investor is dependent on a reference price plus the investor's risk attitudes towards gains and losses.

Given the discrepancy between cumulative prospect theory and risk neutrality, it is expected that agents who behave according to cumulative prospect theory misreport their true beliefs under a proper scoring rule since they are no longer expected value maximizers. In other words, proper scoring rules, as a payment structure, might induce cumulative prospect theory agents to report beliefs other than their true beliefs. This fact happens not necessarily because agents are strategic, but because traditional proper scoring rules do not take agents' risk attitudes into account.

A potential way to deal with the above issue is to calibrate agents' reported beliefs a posteriori. Unfortunately, this is not always possible. We show in this paper that traditional proper scoring rules are naturally incompatible with cumulative prospect theory, in a sense that there might be multiple true beliefs associated with a single reported belief. As a consequence, a reported belief cannot be always mapped back to a unique true belief.

Although this problem was already noted before (Offerman et al. 2009; Kothiyal et al. 2011; Offerman and Palley 2016), previous works focused on the narrow case having probability values around 0.5. We show in this paper that this is not the only case and that, depending on the underlying proper scoring rule, such a problem might happen around any probability value.

We also provide a novel explanation for why a property called *comonotonicity* can solve the above incompatibility problem, and further suggest new strategies to construct comonotonic proper scoring rules from any bounded proper scoring rule. We finally suggest an approach to calibrate an agent's reported belief so as to obtain the agent's true belief when beliefs are elicited by means of a comonotonic proper scoring rule. In short, our proposed method eliminates the influence of the agent's risk attitude, as defined by cumulative prospect theory, on his reported belief.

2 Related work

Several mechanisms for inducing honest reporting of beliefs have been proposed in the literature. We refer the interested readers to the work by Schlag et al. (2015) and Schotter and Trevino (2014) for a detailed review of such mechanisms. Our focus in this paper is on *proper scoring rules* (Winkler and Murphy 1968). Proper scoring rules have been used directly and indirectly to promote honest reporting in a variety of domains, e.g., when grading students' exams (Bickel 2010), when sharing rewards amongst a set of agents based on peer evaluations (Carvalho and Larson 2012), to elicit predictions in patient management and clinical trials (Spiegelhalter 1986), to elicit policy makers' beliefs regarding the occurrence of political and economic events (Tetlock 2005), and in prediction markets set to aggregate agents' subjective probabilities (Hanson 2003; Carvalho 2017).

Proper scoring rules rely on two main assumptions, namely risk neutrality and the existence of observable outcomes. Our focus in this paper is on the first assumption. It is fair to say that risk neutrality is an acceptable assumption for computational agents since one can always argue that such agents are programmable (Parkes and Wellman 2015). When agents are humans, however, the assumption of risk neutrality is often unrealistic, even when the involved stakes are small (Weber and Chapman 2005; Armantier and Treich 2013). Focusing on the quadratic scoring rule, Winkler and Murphy (1970) investigated how nonlinear utility functions affect agents' reporting behavior. For some specific utility functions, Winkler and Murphy (1970) showed that a risk-seeking agent reports a sharp probability distribution, whereas a risk-averse agent reports a distribution close to the uniform distribution. In other words, different utility functions might induce different reporting behavior. Focusing on risk-averse agents, Armantier and Treich (2013) characterized how proper scoring rules might bias reported beliefs for different scoring ranges, when the agent has a financial stake in the event he is predicting, and when the agent can hedge his prediction by taking an additional action whose payoff depends on the outcome of the event.

Winkler and Murphy (1970) and Armantier and Treich (2013) assumed that agents are expected utility maximizers, where different risk attitudes are driven exclusively by utility functions. Modern models of decisions under uncertainty based on rank-dependent utilities suggest that besides nonlinear utility functions, probability sensitivity also plays a role in defining an agent's attitude towards uncertainty (Gilboa 1987; Quiggin 1982; Schmeidler 1989). Carvalho (2015) discussed the consequences of assuming that agents are expected utility maximizers when they actually behave according to rank-dependent utility (RDU) theory and the rank-affected multiplicative weights (RAM) model (Birnbaum 1997, 2008). In the former case, Carvalho (2015) showed that even when a proper scoring rule is tailored to a RDU agent's utility function, that agent still misreports his true belief by reporting a vector of *decision weights*. Decision weights reflect a cognitive bias concerning how human beings deal with probability values when making decisions under risk and uncertainty. Thus, Carvalho's analysis highlighted the importance of knowing all the components that drive an agent's attitude towards uncertainty before appropriately eliciting that agent's belief. The author, however, did not elaborate on why a property called comonotonicity is crucial when eliciting beliefs from RDU agents, which is one of the main arguments in our paper. Moreover, our analysis relies on less stringent technical assumptions.

Offerman et al. (2009) suggested an approach to calibrate probability values reported under the quadratic scoring rule by agents who make decisions based on rank-dependent utilities in settings involving only two outcomes. The focus on a single variant of the quadratic scoring rule makes the method by Offerman et al. (2009) quite limited. Furthermore, the suggested mechanism does not work well when analyzing probability values less than 0.5. Our work generalizes the results by Offerman et al. (2009) in three different ways: (1) by considering multiple outcomes, as opposed to binary outcomes; (2) by considering all comonotonic proper scoring rules, as opposed to only the quadratic scoring rule; and (3) by considering cumulative prospect theory (Tversky and Kahneman 1992), a decision theory based on rank-dependent utilities which also incorporates loss aversion and reference dependence. Cumulative prospect theory has consistently outperformed expected utility theory in terms of predictive accuracy (Camerer 2004; Starmer 2000), thus being a more suitable candidate to describe how agents make decisions under risk and uncertainty.

Kothiyal et al. (2011) noted that traditional proper scoring rules might not actually be proper under rank-dependent models. In particular, focusing on positive versions of the quadratic scoring rule and settings involving binary outcomes, Kothiyal et al. (2011) showed that many subjective probabilities might relate to the reported probability of 0.5 and, hence, such a probability value cannot be used to uniquely determine an agent's true belief. We generalize the results by Kothiyal et al. (2011) in three different ways. First, we show that the above regression problem has nothing to do with the specific probability value of 0.5. Instead, it is dependent on the underlying proper scoring rule. We illustrate how different proper scoring rules might cause similar regression problems around probability values other than 0.5. Second, we provide a novel and more formal explanation for why comonotonicity is a sufficient condition for mapping each reported belief to a single true belief and, consequently, to make proper scoring rules indeed proper under cumulative prospect theory. Finally, we show how to construct a comonotonic proper scoring rule from any bounded proper scoring rule and for any number of outcomes, whereas Kothiyal et al. (2011) focused only on positive proper scoring rules and binary outcomes.

3 Proper scoring rules

We consider a set of exhaustive and mutually exclusive outcomes $\{\theta_1, \ldots, \theta_n\}$, for $n \geq 2$. We are interested in obtaining estimates on the probability of the occurrence of each outcome. To achieve this, we elicit probability vectors over the outcomes from experts, henceforth referred to as *agents*. Agents have no influence on the occurrence of the outcomes, and they have no stakes in the outcomes of interest. We denote an agent's belief by the probability vector $\mathbf{p} = (p_1, \ldots, p_n)$, where p_k is the subjective probability regarding the occurrence of outcome θ_k. Agents are potentially strategic, meaning that they are not necessarily honest when reporting their beliefs. Therefore, we distinguish between an agent's *true belief* $\mathbf{p}$, and his *reported belief* $\mathbf{q} = (q_1, \ldots, q_n)$. Clearly, from a decision making perspective, it is desirable to obtain $\mathbf{q} = \mathbf{p}$. When $\mathbf{q} = \mathbf{p}$, we say that the agent is honestly reporting his belief. We induce honest reporting using payment structures called *proper scoring rules* (Winkler and Murphy 1968). Finally, we assume that each agent behaves according to *cumulative prospect theory* (Tversky and Kahneman 1992) when making decisions under uncertainty. We describe cumulative prospect theory in Sect. 4.

A *scoring rule* $R(\mathbf{q}, \theta_x)$ evaluates the accuracy of a reported belief $\mathbf{q}$ by providing a real-valued score upon observing an outcome θ_x, for $x \in \{1, \ldots, n\}$. Scores are often coupled with relevant incentives, such as financial and/or social-psychological rewards, which implies that agents seek to maximize the obtained scores. A scoring rule is called *strictly proper* when an agent receives his maximum expected score if and only if his reported belief $\mathbf{q}$ corresponds to his true belief $\mathbf{p}$ (Winkler and Murphy 1968). The *expected score* of $\mathbf{q}$ at $\mathbf{p}$ for a real-valued scoring rule $R(\mathbf{q}, \theta_x)$ is:

$$\mathbb{E}_{\mathbf{p}}\left[R(\mathbf{q}, \cdot)\right] = \sum_{x=1}^{n} p_x R(\mathbf{q}, \theta_x)$$

Some of the best known proper scoring rules, together with their scoring ranges are (Carvalho 2016b):

$$\text{spherical:} R(\mathbf{q}, \theta_x) = \frac{q_x}{\sqrt{\sum_{k=1}^{n} q_k^2}} [0, 1] \tag{1}$$

$$\text{logarithmic:} R(\mathbf{q}, \theta_x) = \log q_x (-\infty, 0] \tag{2}$$

$$\text{quadratic:} R(\mathbf{q}, \theta_x) = 2q_x - \sum_{k=1}^{n} q_k^2 [-1, 1] \tag{3}$$

Savage (1971) showed that any differentiable strictly convex function that is well-behaved at the endpoints of the scoring range can generate a proper scoring rule. Schervish (1989) and Gneiting and Raftery (2007) later provided more rigorous versions of the characterization by Savage (1971). An important property of proper scoring rules is that a positive affine transformation of a proper scoring rule is still proper (Toda 1963). Hence, in practice, one can easily change the range of a scoring rule, meaning that the above scoring rules are not always associated with the listed ranges.

Henceforth, we say that a scoring rule is *positive* when all the returned scores are nonnegative, i.e., $R(\mathbf{q}, \theta_x) \geq 0$ for all $x \in \{1, \ldots, n\}$. The spherical scoring rule in (1) is an example of a positive scoring rule. A *negative scoring rule*, on the other hand, only returns nonpositive scores, i.e., $R(\mathbf{q}, \theta_x) \leq 0$ for all $x \in \{1, \ldots, n\}$. The logarithmic scoring rule in (2) is an example of a negative scoring rule. Finally, a *mixed scoring rule* might return both positive and negative scores. The quadratic scoring rule in (3) is an example of a mixed scoring rule.

Throughout this paper, we assume a strict ordering of the outcomes such that the scores from a proper scoring rule are ordered: $R(\mathbf{q}, \theta_n) > R(\mathbf{q}, \theta_{n-1}) > \cdots > R(\mathbf{q}, \theta_1)$ for all $\mathbf{q}$, i.e., an agent will always receive the highest score when outcome θ_n happens, the second highest score when outcome θ_{n-1} happens, and so on, no matter what he reports. In other words, the proper scoring rule R satisfies a property called *comonotonicity* (Schmeidler 1989; Kothiyal et al. 2011). For our purposes, a proper scoring rule R satisfies comonotonicity when $R(\mathbf{q}, \theta_{x+1}) > R(\mathbf{q}, \theta_x)$ for any $x \in \{1, \ldots, n-1\}$ and $\mathbf{q} \in \Delta^{n-1}$, where Δ^{n-1} is the unit simplex in $\mathfrak{R}^n$. We discuss later in this paper why comonotonicity is a crucial property to ensure properness when agents behave according to cumulative prospect theory. We note that the outcomes $\theta_1, \ldots, \theta_n$ can always be rearranged so that the above inequalities hold true.

3.1 Constructing a comonotonic proper scoring rule

We discuss in this subsection how to create comonotonic proper scoring rules from any bounded proper scoring rule by adding stakes to the underlying outcomes, thus extending the work by Kothiyal et al. (2011) who considered proper scoring rules only in domains with binary outcomes. Consistent with the definition of Kadane and Winkler (1988) and Armantier and Treich (2013), we introduce stakes by assuming that scores increase by an exogenous amount when different outcomes occur. For instance, an agent may be asked to predict whether the closing price of a stock he holds will go up in the coming day. The final payoff of that agent is dependent not only on the score from the proper scoring rule, but also on the final stock value, which in turn is dependent on the observed outcome, but independent from the reported prediction. This is formally equivalent to adding different constants to $R(\mathbf{q}, \theta_x)$ for different values of $x \in \{1, \ldots, n\}$. Formally, one can think of stakes as a function that maps outcomes to real values. As a result, it is immediate that any proper scoring rule remains proper in the presence of stakes. In this subsection, we assume that the baseline proper scoring rule R is bounded within the range (a, b). When R is a positive proper scoring rule, i.e., $a, b \in \mathfrak{R}^+$, we can then define a new scoring rule S as follows:

$$S(\mathbf{q}, \theta_x) = R(\mathbf{q}, \theta_x) + (x-1) \cdot (b-a) \tag{4}$$

for $x \in \{1, \ldots, n\}$. The above scoring function has different non-overlapping ranges for different outcomes. When the first outcome happens, i.e., when $x = 1$, the range of the resulting score is (a, b). When θ_2 happens, the range of the resulting score is $(b, 2b-a)$. When θ_3 happens, the range of the resulting score is $(2b-a, 3b-2a)$, and so on. Clearly, we obtain $S(\mathbf{q}, \theta_n) > S(\mathbf{q}, \theta_{n-1}) > \cdots > S(\mathbf{q}, \theta_1) \geq 0$, meaning that S is positive and comonotonic. Moreover, S is trivially proper, i.e., an agent has to report $\mathbf{q} = \mathbf{p}$ to maximize the expected value of S.

When R is a negative proper scoring rule with range (a, b), for $a, b \in \mathfrak{R}^-$, we define S as follows:

$$S(\mathbf{q}, \theta_x) = R(\mathbf{q}, \theta_x) + (n - x) \cdot (a - b) \tag{5}$$

for $x \in \{1, \ldots, n\}$. Once more, S is proper and it has different non-overlapping ranges for different outcomes, thus satisfying comonotonicity. Moreover, we obtain $0 \geq S(\mathbf{q}, \theta_n) > S(\mathbf{q}, \theta_{n-1}) > \cdots > S(\mathbf{q}, \theta_1)$, meaning that S is negative.

Finally, when R is a mixed proper scoring rule satisfying $R(\mathbf{q}, \theta_n) > R(\mathbf{q}, \theta_{n-1}) > \cdots > R(\mathbf{q}, \theta_i) \geq 0 > R(\mathbf{q}, \theta_{i-1}) > \cdots > R(\mathbf{q}, \theta_1)$, bounded within the range (a, b), for $a \in \Re^-$ and $b \in \Re^+$, we can then define a new scoring rule S as follows:

$$S(\mathbf{q}, \theta_x) = R(\mathbf{q}, \theta_x) + (n - x) \cdot (a - b), \quad \text{for } x \in \{1, \ldots, i-1\}$$
$$S(\mathbf{q}, \theta_x) = R(\mathbf{q}, \theta_x) + (x - 1) \cdot (b - a), \quad \text{for} x \in \{i, \ldots, n\}$$

that is, the resulting scoring rule S is a combination of (4) and (5). S is a mixed scoring rule satisfying $S(\mathbf{q}, \theta_n) > S(\mathbf{q}, \theta_{n-1}) > \cdots > S(\mathbf{q}, \theta_i) \geq 0 > S(\mathbf{q}, \theta_{i-1}) > \cdots > S(\mathbf{q}, \theta_1)$, and it is proper and comonotonic. Note that the index i determines when scores move from positive to negative. In particular, when $i = 1$ (respectively, $i = n + 1$), a mixed proper scoring rule becomes a positive (respectively, negative) proper scoring rule.

It is important to note that the the design of a comonotonic proper scoring rule and, consequently, the ranking of the underlying outcomes are done a priori, i.e., before eliciting an agent's belief. Hence, they are independent of any reported belief $\mathbf{q}$. A question that might then arise is: how should a requester order the outcomes? An interesting idea is to minimize each agent's payment when scores are coupled with financial rewards and the requester has a prior/baseline belief over the outcomes. For example, the lowest score range can be associated with the outcome the requester believes is the most likely to occur, the second score range with the second most likely outcome, and so on. Clearly, one can adapt this idea in different ways, e.g., when the requester wants to maximize the collected payments resulting from the use of a negative proper scoring rule. In spirit, our idea is similar to the concept of tailored proper scoring rules by Johnstone et al. (2011), where scoring rules are tailored to specific decision-making problems and to the utility functions of particular requesters. In particular, one can see our suggestion as a way of tailoring a proper scoring rule to a requester's objective regarding scoring rule payments.

To summarize, one can create a comonotonic proper scoring rule from any bounded proper scoring rule by adding different stakes to different outcomes, so that the range of the score an agent receives depends on the observed outcome, where different non-overlapping ranges are associated with different outcomes. Hence, an agent's reported belief defines the agent's scores inside the possible ranges. Our approach is constructive in nature. We note, however, that our approach can be further generalized. In particular, any scheme that defines score ranges that do not intersect will result in comonotonic proper scoring rules. It is noteworthy that none of the traditional proper scoring rules we mentioned in this section satisfies comonotonicity (we elaborate on the quadratic scoring rule case in Sect. 5) and that a positive affine transformation of a comonotonic proper scoring rule is still comonotonic and proper. Moreover, under a comonotonic proper scoring rule, the requester loses some control over the highest/lowest score values assigned to different reported beliefs. This happens because the ranges the scores belong to are now dependent on the observed outcome. For example, one can no longer state that a completely wrong reported belief assigning the probability 1 to an outcome that does not occur will receive a specific score value, say zero, because the range of that score value is now dependent on the underlying observed outcome.

4 Cumulative prospect theory

A basic assumption behind proper scoring rules is that agents behave to maximize their expected scores, i.e., they are risk neutral. It has been widely reported that humans very often fail to be risk neutral, no matter if the underlying stakes are high or low (Starmer 2000; Weber and Chapman 2005). Different from risk neutrality and, more broadly, expected utility theory, modern theories of individual decisions under risk and uncertainty based on *ranks* assert that both sensitivity to payoffs and sensitivity to probabilities generate deviations from risk neutral behavior.

Cumulative Prospect Theory (CPT) (Tversky and Kahneman 1992) is a prominent theory based on ranks which also incorporates loss aversion and reference dependence. CPT is defined in terms of *prospects*, which are outcome-contingent payoffs. We use the notation $\mathbf{f} = [s_1 : \theta_1, \dots, s_n : \theta_n]$ to denote a prospect $\mathbf{f}$ which yields a payoff of $s_k \in \Re$ if outcome θ_k occurs, for $k \in \{1, \dots, n\}$. Since one can always rearrange the outcomes, we assume without loss of generality that $s_n \geq s_{n-1} \geq \cdots \geq s_1$. The uncertainties regarding the occurrence of the outcomes are quantified in terms of a probability vector $\mathbf{p} = (p_1, \dots, p_n)$. Consequently, we can represent a prospect as $\mathbf{f} = [s_1 : p_1, \dots, s_n : p_n]$, which yields a payoff of $s_k \in \Re$ with probability p_k.

A prospect is called *positive* if all payoffs are nonnegative, i.e., $s_n > s_{n-1} \geq \cdots \geq s_1 \geq 0$. We denote a positive prospect by $\mathbf{f}^+$. A prospect is called *negative* if all payoffs are nonpositive, i.e., $0 \geq s_n \geq \cdots \geq s_2 > s_1$. We refer to a negative prospect as $\mathbf{f}^-$. Finally, a *mixed* prospect $\mathbf{f}^\pm$ contains both positive and negative payoffs, i.e., $s_n > s_{n-1} \geq \cdots \geq s_i \geq 0 \geq s_{i-1} \geq \cdots \geq s_2 > s_1$.

In expected utility theory, which subsumes risk neutrality, the utility of a prospect, i.e., the value an agent derives from a prospect, is equal to the sum of the utilities of the payoffs, each one weighted by its underlying (subjective) probability. CPT proposes two major modifications of the expected utility theory framework: (1) the utility of each payoff is multiplied by a *decision weight*, not by an additive probability; and (2) the carriers of value are gains and losses relative to a reference point, not final payoffs. Formally, the value an agent derives from a prospect is defined in terms of a *utility function*, a *loss aversion parameter*, and *weighting functions*. We discuss these components in the following subsections before going into details about CPT utilities.

4.1 Value functions

Cumulative prospect theory asserts that the carriers of value are gains and losses relative to a reference point. When defining positive, negative, and mixed prospects in the beginning of this section, we implicitly assumed that the reference point was equal to 0. Such an assumption will be made throughout the rest of this paper. We argue that this is not a strong assumption since we are dealing with one-shot mechanisms for eliciting beliefs. Moreover, our results are independent of specific values for the reference point. An alternative modeling choice would be to consider the reference point as the current wealth of the reporting agent. We note that such an approach would require the truthful elicitation of wealth, thus overcomplicating the original task of eliciting beliefs.

CPT agents have different risk attitudes towards gains and losses, i.e., with respect to payoffs above and below the reference point. Formally, the intrinsic value of a payoff s_k is defined in terms of a strictly increasing function $V : \Re \rightarrow \Re$, called the *value function*, which satisfies:

$$V(s_k) = U(s_k) \quad \text{for } s_k \geq 0$$
$$V(s_k) = \lambda U(s_k) \quad \text{for } s_k < 0$$

where $U : \Re \rightarrow \Re$ is a continuously differentiable, strictly increasing utility function satisfying $U(s_k) \geq 0$, for $s_k \geq 0$, and $U(s_k) \leq 0$, for $s_k \leq 0$. The parameter $\lambda \geq 1$ is the *loss aversion* parameter. The loss aversion parameter captures the psychological phenomenon that "losses loom larger than gains" (Tversky and Kahneman 1992). As a consequence, the value function is steeper for losses than for gains, i.e., $V'(s_k) < V'(-s_k)$, for $s_k > 0$. It is often the case that the parameter λ is defined in terms of the utility function. For example, Tversky and Kahneman (1992) implicitly assumed that $\lambda = -\frac{U(-1)}{U(1)}$, finding empirically that $\lambda = -U(-1) = 2.25$. Another property of the value function V is that it is concave for gains, convex for losses, and it satisfies $V(0) = 0$. Tversky and Kahneman (1992) proposed the following value function:

$$V(s_k) = \begin{cases} s_k^{\alpha} & \text{for } s_k \geq 0, \\ \lambda \cdot \left(-(-s_k)^{\beta}\right) & \text{otherwise} \end{cases} \qquad (6)$$

where α and β are parameters of the underlying power utility function. Figure 1 illustrates the value function in (6) for parameter values empirically determined by Tversky and Kahneman (1992).

4.2 Weighting functions

Cumulative prospect theory asserts that the weight associated with the value $V(s_k)$ is the difference between two transformed *ranks*, instead of an individual probability value p_k, as in expected utility theory. Following the notation in the book by Wakker (2010), for a positive prospect $\mathbf{f}^+$, the rank of a payoff s_k is the probability of $\mathbf{f}^+$ yielding a payoff better than s_k, i.e., the rank of s_k is equal to $p_{k+1} + p_{k+2} + \cdots + p_n$. In this way, ranks are numbers between 0 and 1 and not integers. The weight of $V(s_k)$ is then the marginal contribution of the individual probability p_k to the total probability of

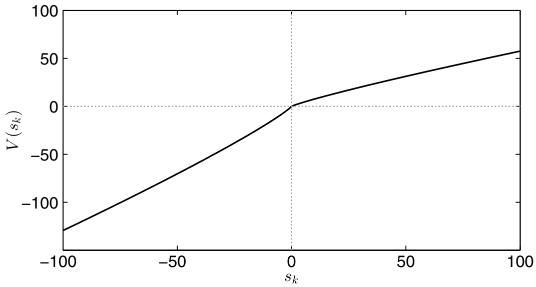


Fig. 1 Tversky and Kahneman's value function for $\alpha = \beta = 0.88$, $\lambda = 2.25$, and the reference point equal to 0

receiving better payoffs, measured in terms of a continuous and strictly increasing function $W^+ : [0,1] \to [0,1]$. Formally, the weight of $V(s_k)$ is π_k^+, where:

$$\pi_n^+ = W^+(p_n)$$
$$\pi_k^+ = W^+\left(\sum_{x=k}^{n} p_x\right) - W^+\left(\sum_{x=k+1}^{n} p_x\right), \quad \text{for } k \in \{1, \ldots, n-1\} \tag{7}$$

For a negative prospect, the weight associated with $V(s_k)$, π_k^-, is the marginal contribution of the individual probability p_k to the total probability of receiving worse payoffs, measured in terms of a continuous and strictly increasing function $W^- : [0,1] \to [0,1]$, i.e.:

$$\pi_1^- = W^-(p_1)$$
$$\pi_k^- = W^-\left(\sum_{x=1}^{k} p_x\right) - W^-\left(\sum_{x=1}^{k-1} p_x\right), \quad \text{for } k \in \{2, \ldots, n\} \tag{8}$$

The π-values in (7) and (8) are traditionally referred to as *decision weights*. Decision weights reflect a cognitive bias concerning how human beings distort probability values and, thus, they should not be taken as a measure of an agent's true belief. Importantly, $(\pi_1^+, \ldots, \pi_n^+)$ and $(\pi_1^-, \ldots, \pi_n^-)$ are probability vectors because $W^+\left(\sum_{k=1}^n p_k\right) = W^-\left(\sum_{k=1}^n p_k\right) = 1$, whereas the same is only true for $(\pi_1^-, \ldots, \pi_{i-1}^-, \pi_i^+, \ldots, \pi_n^+)$ when $W^-(\rho) + W^+(1-\rho) = 1$, for all $\rho \in [0,1]$.

The functionals W^+ and W^- are known as *weighting functions*. Common findings suggest that a weighting function is a nonlinear transformation of the probability scale that overweights small probabilities and underweights moderate and high probabilities (Abdellaoui 2000; Tversky and Kahneman 1992). In other words, the weighting function displays an inverse-S shape: it is concave near 0 and convex near 1. Moreover, $W^+(0) = W^-(0) = 0$. The weighting functions proposed by Tversky and Kahneman (1992) for gains and loses are:

$$W^+(\rho) = \frac{\rho^\gamma}{(\rho^\gamma + (1-\rho)^\gamma)^{\frac{1}{\gamma}}} \quad \text{and} \quad W^-(\rho) = \frac{\rho^\delta}{(\rho^\delta + (1-\rho)^\delta)^{\frac{1}{\delta}}} \tag{9}$$

where $\gamma, \delta \geq 0.28$ for the weighting functions to be strictly increasing. Figure 2 illustrates the above weighting functions for $\gamma = 0.61$ and $\delta = 0.69$, the parameter values empirically found by Tversky and Kahneman (1992). We refer the interested reader to the work by Gonzalez and Wu (1999) for a review of different shapes of weighting functions.

4.3 CPT utility

Cumulative prospect theory asserts that risk attitudes are jointly determined by the weighting functions and the value function. Together, they result in a fourfold pattern of risk attitudes: risk aversion for gains and risk seeking for losses of high probability; risk seeking for gains and risk aversion for losses of low probability (Tversky and Kahneman 1992). Let CPT($\mathbf{f}$) be the value an agent derives from a prospect according to cumulative prospect theory. For a positive prospect $\mathbf{f}^+ = [s_1 : p_1, \ldots, s_n : p_n]$, CPT($\mathbf{f}^+$) is defined as:

$$\mathrm{CPT}(\mathbf{f}^+) = \sum_{k=1}^{n} \pi_k^+ V(s_k) \tag{10}$$

The CPT utility for a negative prospect $\mathbf{f}^- = [s_1 : p_1, \ldots, s_n : p_n]$ is:

$$\mathrm{CPT}(\mathbf{f}^-) = \sum_{k=1}^{n} \pi_k^- V(s_k) \tag{11}$$

Finally, the CPT utility for a mixed prospect $\mathbf{f}^\pm = [s_1 : p_1, \ldots, s_n : p_n]$, where $s_n > s_{n-1} \geq \cdots \geq s_i \geq 0 \geq s_{i-1} \geq \cdots \geq s_2 > s_1$, is:

$$\mathrm{CPT}\left(\mathbf{f}^\pm\right) = \sum_{k=1}^{i-1} \pi_k^- V(s_k) + \sum_{k=i}^{n} \pi_k^+ V(s_k) \tag{12}$$

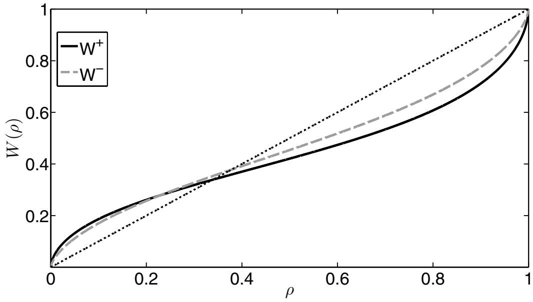


Fig. 2 Tversky and Kahneman's weighting functions for $\gamma = 0.61$ and $\delta = 0.69$

4.4 Numerical example

The following example illustrates how risk neutral and CPT agents might behave differently when evaluating prospects. Consider an agent facing the choice between two positive prospects, namely $\mathbf{f}_1 = [150, 1]$ and $\mathbf{f}_2 = [10 : 0.6, 100 : 0.3, 1000 : 0.1]$. In words, the former prospect pays 150 units of numeraire for sure, whereas the latter prospect has a relatively high (respectively, low) probability of paying less (respectively, more) than the former prospect. The value of prospect $\mathbf{f}_1$ to a risk neutral agent is equal to $1 \times 150 = 150$, whereas the value of prospect $\mathbf{f}_2$ is equal to $10 \times 0.6 + 100 \times 0.3 + 1000 \times 0.1 = 136$. Hence, a risk neutral agent prefers prospect $\mathbf{f}_1$ over prospect $\mathbf{f}_2$.

Now, let us analyze how a CPT agent evaluates the above prospects. We assume the utility and weighting functions represented by, respectively, Eqs. 6 and 9, the parameter values being equal to the ones in Figs. 1 and 2. The CPT agent evaluates prospect $\mathbf{f}_1$ as follows: $\text{CPT}(\mathbf{f}_1) = W^+(1)V(150) = 150^{0.88} = 82.21675$. Note how the value function affects the valuation of the prospect, whereas the weighting function does not. This happens because the prospect is deterministic, i.e, there is no uncertainty involved in its payment. Now, consider how the CPT agent evaluates prospect $\mathbf{f}_2$: $\text{CPT}(\mathbf{f}_2) = W^+(0.1)V(1000) + (W^+(0.3 + 0.1) - W^+(0.1))V(100) + (1 - W^+(0.3+ 0.1))V(10) = 96.6749$. That is, the CPT agent prefers gambling with $\mathbf{f}_2$ as opposed to the deterministic prospect $\mathbf{f}_1$.

One can also consider the intermediate case where agents are expected utility maximizers. The value of prospect $\mathbf{f}_1$ to that agent is equal to $1 \times V(150) = 150^{0.88} = 82.2167$, whereas the value of prospect $\mathbf{f}_2$ is equal to $V(10) \times 0.6 + V(100) \times 0.3 + V(1000) \times 0.1 = 65.4662$. Hence, a risk-averse agent prefers prospect $\mathbf{f}_1$ over prospect $\mathbf{f}_2$.

This example illustrates some interesting points. First, it shows how different decision models might imply that the underlying agents prefer different prospects. Second, it highlights that risk aversion in the CPT model is defined by both weighting functions and value functions, whereas the risk attitude of expected utility maximizers is driven exclusively by the value function.

5 Proper scoring rules under cumulative prospect theory

We note that the scores resulting from a comonotonic proper scoring rule R can be stated in terms of a prospect, i.e., $[R(\mathbf{q}, \theta_1) : p_1, \dots, R(\mathbf{q}, \theta_n) : p_n]$. In other words, when reporting a belief $\mathbf{q}$, an agent is essentially defining the payoffs of a prospect, where the associated probabilities are subjective probabilities. Reporting a belief $\mathbf{q}$ is then equivalent to making a decision under uncertainty by choosing a prospect among a potentially infinite number of prospects. Consequently, an agent's reporting behavior can be analyzed from the perspective of different decision theories, including cumulative prospect theory.

In general, the utility derived by an agent who behaves according to cumulative prospect theory from a prospect is only equal to the utility derived by a risk neutral agent when: (1) the weighting functions are identity functions; (2) the utility function is linear; and (3) the loss aversion parameter is equal to 1. Hence, cumulative prospect theory subsumes risk neutrality and, more broadly, expected utility theory. As discussed before, common findings suggest that the weighting functions are nonlinear functions, the utility function is concave for gains and convex for losses, and the loss aversion parameter is greater than 1. Thus, risk neutral agents and CPT agents found in practice are expected to value prospects in different ways and, consequently, to behave differently when reporting their beliefs under a proper scoring rule.

Later in this paper, we characterize how CPT agents report their beliefs under positive, negative, and mixed proper scoring rules. Moreover, we discuss how to solve the resulting systems of equations to obtain an agent's true belief from his reported belief. Naturally, we start by assuming that an agent is rational, meaning that he reports his belief so as to maximize his CPT utility. In other words, each agent is solving the following optimization problem:

$$\begin{aligned} &\underset{\mathbf{q}}{\text{maximize}}\,\text{CPT}(\mathbf{q}) \\ &\text{subject to} \sum_{k=1}^{n} q_k = 1 \\ &\qquad\qquad q_k \geq 0, k \in \{1, \dots, n\} \end{aligned} \tag{13}$$

One might argue that cumulative prospect theory is a descriptive theory, as opposed to a prescriptive, objective-maximizing theory. We note, however, that our approach is in agreement with relevant literature, e.g., see the work by Offerman et al. (2009), Kothiyal et al. (2011), and Offerman and Palley (2016). We abuse notation by dropping the underlying proper scoring rule R and by writing $\text{CPT}(\mathbf{q})$, instead of $\text{CPT}(\mathbf{f})$, for $\mathbf{f} = [R(\mathbf{q}, \theta_1) : p_1, \dots, R(\mathbf{q}, \theta_n) : p_n]$. Given that the CPT functional is continuous and that the domain of every q_k is compact, an optimal $\mathbf{q}$ always exists. There may exist several optima, in which case one optimum is arbitrarily selected to be the reported probability vector $\mathbf{q}$. We note that, since all the constraints are linear, it must be the case that the reported probability vector $\mathbf{q}$ satisfies the Karush–Kuhn–Tucker (KKT) conditions, i.e., it satisfies first-order necessary conditions for a solution in a nonlinear optimization problem to be optimal. The Lagrangian associated with the above problem is:

$$L(\mathbf{q}, \tau, \boldsymbol{\mu}) = \text{CPT}(\mathbf{q}) - \tau\left(\sum_{k=1}^{n} q_k - 1\right) + \sum_{k=1}^{n} \mu_k q_k \tag{14}$$

One can think of the above Lagrangian as a reformulation of the optimization problem in (13) that penalizes violations of its inequality constraints. In our upcoming analyses, we heavily use the Lagrangian in a reverse-engineering fashion, i.e., knowing what the optimal solution to the above optimization problem is (namely, $\mathbf{q}$), we will find the reported belief $\mathbf{p}$ that leads to that solution. The dual feasibility and the complementary slackness associated with (14) are:

$$\mu_k \geq 0, \quad \text{for } k \in \{1, \ldots, n\}$$
$$\mu_k q_k = 0, \quad \text{for } k \in \{1, \ldots, n\}$$

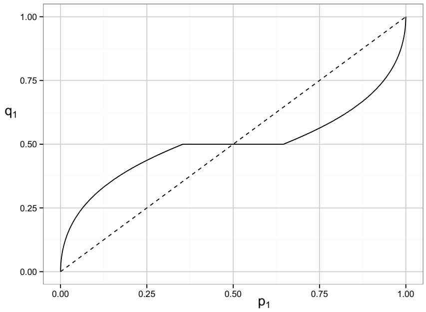
The dual feasibility and complementary slackness are part of the KKT conditions, meaning that they must be satisfied for a reported belief $\mathbf{q}$ to be an optimal solution to the problem in (13). Let $\mathbf{q}^* = (q_1^*, \ldots, q_n^*), \tau^*, \boldsymbol{\mu}^* = (\mu_1^*, \ldots, \mu_n^*)$ be the optimal points for the Lagrangian. Henceforth, we assume that the constraint $q_k \geq 0$ is not binding at the optimal, i.e., $q_k^* > 0$, for $k \in \{1, \ldots, n\}$. Consequently, due to the complementary slackness, $\mu_k^* = 0$, for $k \in \{1, \ldots, n\}$. This assumption allows us to remove $\boldsymbol{\mu}$ from the Lagrangian. We note that without such an assumption, we might run into technical problems whenever the underlying proper scoring rule is not bounded, e.g., the resulting CPT utility is undefined when the proper scoring rule is the logarithmic scoring rule and $q_k^* = 0$, for some $k \in \{1, \ldots, n\}$. We allow, however, q_k^* to be arbitrarily close to 0.

5.1 Why comonotonicity matters

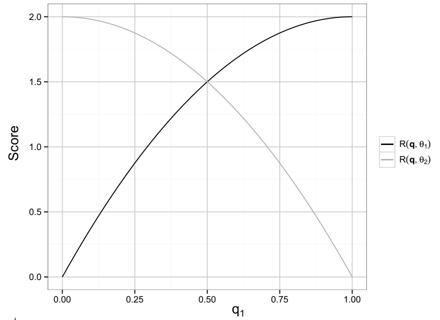
A question that might arise from the above discussion is: why does the proper scoring rule have to satisfy comonotonicity? Under a non-comonotonic proper scoring rule, agents with cumulative prospect theory utilities and different true beliefs might end up reporting a similar belief, which means that the reported belief cannot be used to uniquely identify an agent's true belief. This point was first noted by Offerman et al. (2009) for a positive version of the quadratic scoring rule and binary outcomes. For illustration's sake, consider an agent with belief $\mathbf{p} = (p_1, p_2) = (p_1, 1 - p_1)$, who reports $\mathbf{q} = (q_1, q_2) = (q_1, 1 - q_1)$ under the proper scoring rule $R(\mathbf{q}, \theta_x) = 2q_x - \left(\sum_{k=1}^{n} q_k^2\right) + 1$, which is positive and has the range [0, 2]. Figure 3a illustrates the findings by Offerman et al. (2009), i.e., how an agent with the CPT utility in (10) reports q_1 in terms of p_1 when W^+ is equal to the weighting function in (9) with parameter $\gamma = 0.61$, and the value function is equal to the function in (6) with parameter $\alpha = 0.88$.

The most striking feature of Fig. 3a is that the curve is flat around $p_1 = 0.5$. As suggested by Offerman et al. (2009), the risk aversion generated by the CPT utility is so strong for a subjective probability around 0.5 that an agent makes the safe choice of reporting 0.5. Putting it in different words, Fig. 3a shows that the underlying non-comonotonic proper scoring rule may be insensitive to small changes in the neighborhood of 0.5 for binary outcomes. As a consequence, a reported probability of $q_1 = 0.5$ relates to many degrees of true belief and, thus, it cannot be used to uniquely determine the correct true belief. As Kothiyal et al. (2011) mentioned, if properness is to be taken to mean that all degrees of belief can be identified, then traditional, non-comonotonic proper scoring rules are no longer proper under cumulative prospect theory.

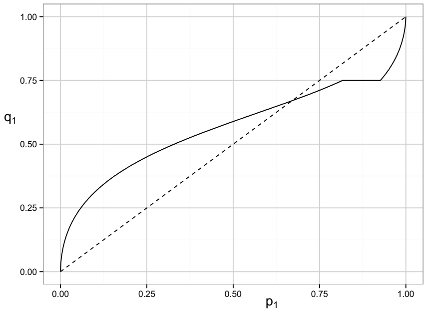
When discussing the flat-region issue, previous work focused on the quadratic scoring rule, the specific probability value of 0.5, and binary outcomes (Offerman et al. 2009; Kothiyal et al. 2011; Offerman and Palley 2016). We argue, however, that this narrow focus is not exhaustive, and that the flat-region issue has nothing



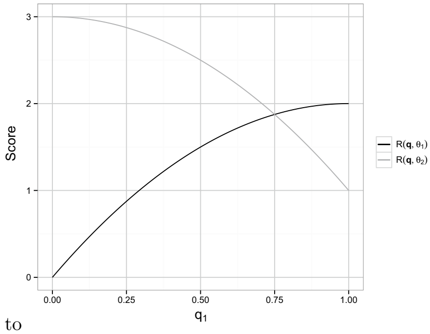
(a) The reported probability of $q_1 = 0.5$ relates to many degrees of true belief.



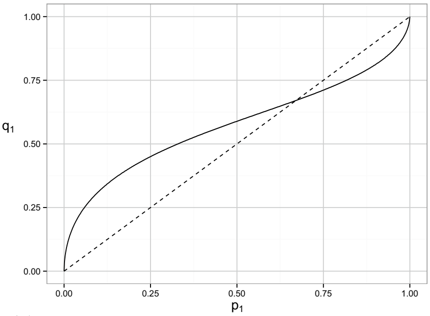
(b) Scores crossing at $q_1 = 0.5$.



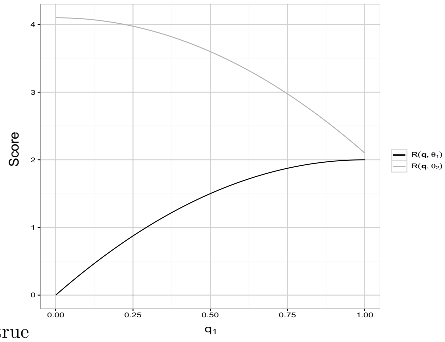
(c) The reported probability of $q_1 = 0.75$ relates to many degrees of true belief.



(d) Scores crossing at $q_1 = 0.75$.



(e) Each reported belief is associated with a single true belief.



(f) No crossing point in the range of $q_1 \in [0, 1]$.

Fig. 3 Example of the impact of comonotonicity on how reported beliefs associate with true beliefs

to do with the specific probability value of 0.5 per se. This is just an artifact of the underlying proper scoring rule, namely the quadratic scoring rule. Instead, we argue that this issue arises at the crossing point when $R(\mathbf{q}, \theta_1) = R(\mathbf{q}, \theta_2)$, for a non-comonotonic proper scoring rule R. It is well known that for binary outcomes, $R(\mathbf{q}, \theta_1)$ is increasing in q_1, whereas $R(\mathbf{q}, \theta_2)$ must be decreasing. This implies that, for the quadratic scoring rule, $R(\mathbf{q}, \theta_1)$ and $R(\mathbf{q}, \theta_2)$ cross once. In the above example, this happens when $q_1 = 0.5$ (see Fig. 3b). To understand the consequences of such a crossing point, consider the case where a CPT agent contemplates reporting the belief $\mathbf{q}' = (0.5, 0.5)$. The CPT utility value assigned to the underlying prospect is then:

$$\mathrm{CPT}(\mathbf{q}') = W^+(p_1)V(R(\mathbf{q}', \theta_1)) + \big(1 - W^+(p_1)\big)V(R(\mathbf{q}', \theta_2)) = V(R(\mathbf{q}', \theta_1)) = V(R(\mathbf{q}', \theta_2)) \tag{15}$$

Note that the CPT value does not depend on the agent's true belief $\mathbf{p}$, and the reason for this is that $R(\mathbf{q}', \theta_1) = R(\mathbf{q}', \theta_2)$. Now, let's contrast the above to the case where the agent contemplates reporting $\mathbf{q}''$, which has $q_1 > 0.5$. This implies that $R(\mathbf{q}'', \theta_1) > R(\mathbf{q}', \theta_1) = R(\mathbf{q}', \theta_2) > R(\mathbf{q}'', \theta_2)$. The CPT value assigned to the underlying prospect is then:

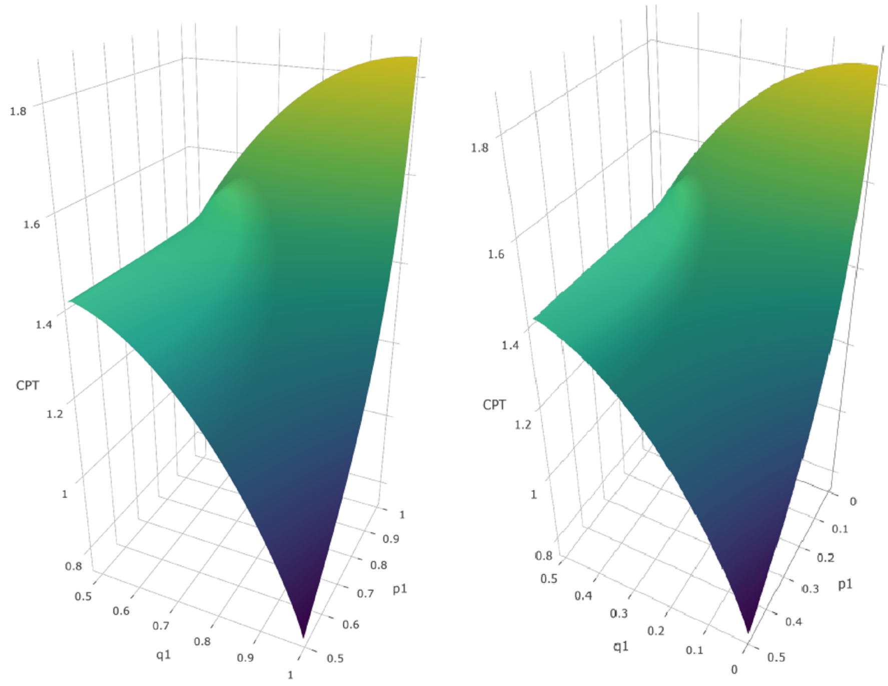
$$\begin{aligned}\mathrm{CPT}(\mathbf{q}'') &= W^+(p_1)V\big(R(\mathbf{q}'', \theta_1)\big) + \big(1 - W^+(p_1)\big)V\big(R(\mathbf{q}'', \theta_2)\big)\\ &= W^+(p_1)\big(V(R(\mathbf{q}'', \theta_1)) - V(R(\mathbf{q}'', \theta_2))\big) + V(R(\mathbf{q}'', \theta_2))\\ &= \epsilon_1 + V(R(\mathbf{q}'', \theta_2))\end{aligned} \tag{16}$$

Let us momentarily disregard the case where a CPT agent's true belief contains $p_1 < 0.5$ (i.e., the leftmost part of Fig. 3a). The question that arises is then: should that agent report the belief $\mathbf{q}'$ or another belief $\mathbf{q}''$? Clearly, the answer to this question depends on the value of ϵ_1 in (16), which in turn depends on the agent's true belief $\mathbf{p} = (p_1, 1 - p_1)$. Recall that $R(\mathbf{q}'', \theta_2) < R(\mathbf{q}', \theta_2)$ and, consequently, $V(R(\mathbf{q}'', \theta_2)) < V(R(\mathbf{q}', \theta_2))$. Equations (15) and (16) then imply that for some values of p_1 and, consequently, ϵ_1, a CPT agent might be better off by reporting $\mathbf{q}'$ as opposed to $\mathbf{q}''$. In the scenario described in Fig. 3a, this is true for $p_1 \in [0.5, 0.646]$, i.e., it is in the best interest of a CPT agent to report $\mathbf{q}' = (0.5, 0.5)$ when $p_1 \in [0.5, 0.646]$, and to report $\mathbf{q}''$ having $q_1 > 0.5$ when $p_1 \in (0.646, 1]$. One can see the resulting CPT values for different reporting strategies in Fig. 4a.

Let's now contrast (15) to the case where the agent contemplates reporting $\mathbf{q}'''$, which has $q_1 < 0.5$. This implies that $R(\mathbf{q}''', \theta_1) < R(\mathbf{q}', \theta_1) = R(\mathbf{q}', \theta_2) < R(\mathbf{q}''', \theta_2)$. The CPT utility value assigned to the underlying prospect is then:

$$\begin{aligned}\mathrm{CPT}(\mathbf{q}''') &= \big(1 - W^+(1 - p_1)\big)V\big(R(\mathbf{q}''', \theta_1)\big) + W^+(1 - p_1)V\big(R(\mathbf{q}''', \theta_2)\big)\\ &= W^+(1 - p_1)(V(R(\mathbf{q}''', \theta_2)) - V(R(\mathbf{q}''', \theta_1))) + V(R(\mathbf{q}''', \theta_1))\\ &= \epsilon_2 + V(R(\mathbf{q}''', \theta_1))\end{aligned} \tag{17}$$

This time, we momentarily disregard the case where a CPT agent's true belief contains $p_1 > 0.5$ (i.e., the rightmost part of Fig. 3a). Whether an agent reports the belief $\mathbf{q}'$ or another belief $\mathbf{q}'''$ depends on the value of ϵ_2 in (17), which in turn depends on the agent's true belief $\mathbf{p} = (p_1, 1 - p_1)$. Recall that $R(\mathbf{q}''', \theta_1) < R(\mathbf{q}', \theta_1)$ and, consequently, $V(R(\mathbf{q}''', \theta_1)) < V(R(\mathbf{q}', \theta_1))$. Equations (15) and (17) then imply that for some values of p_1 and, consequently, ϵ_2, a CPT agent might be better off by reporting $\mathbf{q}'$ as opposed to $\mathbf{q}'''$. In the scenario described in Fig. 3a, this is true for $p_1 \in [0.354, 0.5]$, i.e., it is in the best interest of a CPT agent to report $\mathbf{q}' = (0.5, 0.5)$



(a) $0.5 < p_1, q_1 \leq 1$ **(b)** $0 \leq p_1, q_1 < 0.5$

Fig. 4 CPT values for different true and reported beliefs

when $p_1 \in [0.354, 0.5]$, and to report $\mathbf{q}'''$ having $q_1 < 0.5$ when $p_1 \in [0, 0.354)$. One can see the resulting CPT values for different reporting strategies in Fig. 4b.

Putting the above results together, CPT agents with beliefs containing $p_1 \in [0.354, 0.646]$ are better off by reporting $q_1 = 0.5$ than any other probability value. This explains the flat region in Fig. 3a. In different words, when $q_1 = 0.5$, the ordering of the scores from the quadratic scoring rule for the two outcomes changes and, as a consequence, the objective function changes due to the fact that decision weights change in a drastic non-smooth manner. Mathematically, the abovementioned problem occurs when a proper scoring rule R allows for $R(\mathbf{q}, \theta_x) = R(\mathbf{q}, \theta_y)$, for some probability vector $\mathbf{q}$, and $x, y \in \{1, \ldots, n\}$, i.e., when the proper scoring rule is non-comonotonic. To illustrate that this flat-region problem is not specific to the probability value of 0.5, consider now a non-comonotonic proper scoring rule defined as follows:

$$R(\mathbf{q}, \theta_1) = 2q_1 - q_1^2 - (1 - q_1)^2 + 1$$
$$R(\mathbf{q}, \theta_2) = 2 \cdot (1 - q_1) - q_1^2 - (1 - q_1)^2 + 2$$

that is, the above proper scoring rule is equal to the quadratic scoring rule, but with an extra unit added to the score when outcome θ_2 occurs. When $R(\mathbf{q}, \theta_1) = R(\mathbf{q}, \theta_2)$, the crossing point becomes $q_1 = 0.75$. (see Fig. 3d). Consequently, the reported belief of $q_1 = 0.75$ relates to many true beliefs inside the range [0.816, 0.926] (see

Fig. 3c). An interesting follow-up question is then: what happens when using the following proper scoring rule:

$$\begin{aligned} R(\mathbf{q}, \theta_1) =& 2q_1 - q_1^2 - (1-q_1)^2 + 1 \\ R(\mathbf{q}, \theta_2) =& 2 \cdot (1-q_1) - q_1^2 - (1-q_1)^2 + 3.1 \end{aligned} \tag{18}$$

The above proper scoring rule is equal to the quadratic scoring rule, but with 2.1 extra units added to the score when outcome θ_2 occurs. Consequently, no matter the reported belief, an agent will always receive a score when outcome θ_2 occurs greater than the score when θ_1 occurs. i.e., comonotonicity is satisfied. In this case, there is no longer a crossing point in the range of $q_1 \in [0, 1]$ (see Fig. 3f) and, thus, all reported beliefs correspond to a single true belief, i.e., there is no longer a flat region (see Fig. 3e).

Focusing on positive proper scoring rules and binary outcomes, Kothiyal et al. (2011) suggested that one can avoid the flat region around 0.5 using a comonotonic proper scoring rule. In the following propositions, we provide a more formal and complete explanation on why comonotonicity eliminates any flat region by mapping each reported belief to a single true belief. In particular, we show that this happens not only for binary outcomes and positive proper scoring rules, but for any number of outcomes as well as positive, negative, and mixed proper scoring rules. In other words, we provide a sufficient condition under which proper scoring rules might indeed become proper under cumulative prospect theory.

In the following proposition, recall that a reported belief **q** is optimal, in a sense that it is a solution to the optimization problem in (13). Clearly, that belief has an associated CPT utility value, which is a maximum given **p**, W, and V. We then investigate which true belief **p** results in that maximum value. Our result below says that the maximum value relates to a single true belief **p**, which in turn implies that the associated reported belief **q** also relates to a single true belief **p**.

Proposition 1 *If a mixed proper scoring rule R satisfies comonotonicity, then one can relate each reported belief* **q** *to a single true belief* **p**.

Proof Consider a mixed proper scoring rule R that satisfies comonotonicity, i.e., $R(\mathbf{q}, \theta_n) > R(\mathbf{q}, \theta_{n-1}) > \cdots > R(\mathbf{q}, \theta_i) \geq 0 > R(\mathbf{q}, \theta_{i-1}) > \cdots > R(\mathbf{q}, \theta_1)$. An agent's CPT utility in (12) can then be written as:

$$\begin{aligned} \mathrm{CPT}(\mathbf{q}) = & W^+(p_n)[V(R(\mathbf{q}, \theta_n)) - V(R(\mathbf{q}, \theta_{n-1}))] \\ & + \cdots + W^+\left(\sum_{x=i+1}^{n} p_x\right)[V(R(\mathbf{q}, \theta_{i+1})) - V(R(\mathbf{q}, \theta_i))] \\ & + W^+\left(\sum_{x=i}^{n} p_x\right)V(R(\mathbf{q}, \theta_i)) + W^-\left(\sum_{x=1}^{i-1} p_x\right)V(R(\mathbf{q}, \theta_{i-1})) \\ & + W^-\left(\sum_{x=1}^{i-2} p_x\right)[V(R(\mathbf{q}, \theta_{i-2})) - V(R(\mathbf{q}, \theta_{i-1}))] \\ & + \cdots + W^-(p_1)[V(R(\mathbf{q}, \theta_1)) - V(R(\mathbf{q}, \theta_2))] \end{aligned}$$

Note that **q** is known since it is the agent's reported belief. Moreover, CPT(**q**) is the maximum CPT utility value given W, V, and **p**. We now turn ourselves to the question: which true beliefs **p** can produce CPT(**q**)? To answer this question, we differentiate the above function with respect to a probability value p_k, first for some $k \in \{i, \dots, n\}$, to obtain:

$$\begin{aligned} &\frac{\partial W^+\left(\sum_{x=k}^n p_x\right)}{\partial p_k}[V(R(\mathbf{q}, \theta_k)) - V(R(\mathbf{q}, \theta_{k-1}))] \\ &\quad + \dots + \frac{\partial W^+\left(\sum_{x=i+1}^n p_x\right)}{\partial p_k}[V(R(\mathbf{q}, \theta_{i+1})) - V(R(\mathbf{q}, \theta_i))] \\ &\quad + \frac{\partial W^+\left(\sum_{x=i}^n p_x\right)}{\partial p_k} V(R(\mathbf{q}, \theta_i)) \end{aligned} \tag{19}$$

Since W^+ is strictly increasing, $\frac{\partial W^+(\cdot)}{\partial p_k}$ is greater than zero. Moreover, due to comonotonicity, $R(\mathbf{q}, \theta_n) > R(\mathbf{q}, \theta_{n-1}) > \dots > R(\mathbf{q}, \theta_i) \geq 0$. Since V is strictly increasing, then $V(R(\mathbf{q}, \theta_x)) - V(R(\mathbf{q}, \theta_{x-1})) > 0$, for any $x \in \{i, \dots, n\}$. Hence, the first derivative shown above is greater than zero, which implies that the CPT utility is strictly increasing in p_k when the other probabilities as well as the reported belief **q** are fixed. If, on the other hand, $k \in \{1, \dots, i-1\}$, we obtain:

$$\begin{aligned} &\frac{\partial W^-\left(\sum_{x=1}^{i-1} p_x\right)}{\partial p_k} V(R(\mathbf{q}, \theta_{i-1})) \\ &\quad + \frac{\partial W^-\left(\sum_{x=1}^{i-2} p_x\right)}{\partial p_k}[V(R(\mathbf{q}, \theta_{i-2})) - V(R(\mathbf{q}, \theta_{i-1}))] \\ &\quad + \dots + \frac{\partial W^-\left(\sum_{x=1}^{k} p_x\right)}{\partial p_k}[V(R(\mathbf{q}, \theta_k)) - V(R(\mathbf{q}, \theta_{k+1}))] \end{aligned} \tag{20}$$

Since W^- is strictly increasing, $\frac{\partial W^-(\cdot)}{\partial p_k}$ is greater than zero. Moreover, since $0 > R(\mathbf{q}, \theta_{i-1}) > R(\mathbf{q}, \theta_{i-2}) > \dots > R(\mathbf{q}, \theta_1)$ because of comonotonicity and V is strictly increasing, then $V(R(\mathbf{q}, \theta_x)) - V(R(\mathbf{q}, \theta_{x+1})) < 0$, for any $x \in \{1, \dots, i-1\}$. Hence, the first derivative shown above is strictly less than zero.

As a consequence of the first derivatives never being equal to zero, an agent's CPT utility either increases or decreases with each individual p_k given a fixed **q**, for all $k \in \{1, \ldots, n\}$. This means that the optimal CPT utility value is associated with a specific p_k value, and this is true for all $k \in \{1, \ldots, n\}$. As a consequence, the reported belief $\mathbf{q}$, which produces the optimal CPT utility value, relates to a single $\mathbf{p}$. Recall that we previously mentioned in this section that there may exist several optima to the optimization problem, in which case one optimum is arbitrarily selected to be the reported probability vector $\mathbf{q}$. Since all these potential optima result in the same maximum CPT utility value, they are all then related to the same true belief $\mathbf{p}$. □

Proposition 1 shows that for any given value of $\mathbf{q}$, there is only one value of $\mathbf{p}$, thus eliminating any CPT confounds that may exist in the observed $\mathbf{q}$ values. Note that if comonotonicity was not true in the above proof, then all the subtractions inside the square brackets in (19) and (20) could potentially be equal to zero. This would result in the first derivatives being equal to zero and, consequently, the existence of stationary points, i.e., a single CPT utility value and the underlying reported belief being associated with many true beliefs. This is precisely the case described in Figs. 3a, c.

Corollary 1 *If a positive proper scoring rule R satisfies comonotonicity, then one can relate each reported belief* $\mathbf{q}$ *to a single true belief* $\mathbf{p}$.

Corollary 2 *If a negative proper scoring rule R satisfies comonotonicity, then one can relate each reported belief* $\mathbf{q}$ *to a single true belief* $\mathbf{p}$.

The proofs of Corollaries 1 and 2 follow immediately from the proof of Proposition 1 by considering only the positive and negative scores.

6 Obtaining true beliefs from reported beliefs

In the previous section, we have established that comonotonicity is a crucial property to map each reported belief back to a single true belief. However, we have not discussed so far how to obtain a CPT agent's true belief from his reported belief when the elicitation is performed by means of a comonotonic proper scoring rule. In this section, we suggest procedures to calibrate reported beliefs by removing the influence of value functions and weighting functions to obtain an agent's true belief.

In particular, we derive systems of equations in terms of an agent's value function, weighting functions, true belief, and reported belief that, by construction, must be valid. We further discuss how to solve these systems of equations to obtain an agent's true belief. In our upcoming analysis, we implicitly assume that the components of the CPT utility are known, i.e., the utility function, the loss aversion parameter, and the weighting functions were previously elicited through some elicitation procedure [e.g., see the work by Abdellaoui (2000), Abdellaoui et al. (2008), Hines and Larson (2010), Perny et al. (2016), Wakker and Deneffe (1996)]. We elaborate on why this assumption is necessary in Sect. 7.

6.1 Positive and comonotonic proper scoring rules

Suppose that an agent's belief is elicited through a positive, comonotonic proper scoring rule R, for $R(\mathbf{q}, \theta_n) > R(\mathbf{q}, \theta_{n-1}) > \cdots > R(\mathbf{q}, \theta_1) \geq 0$. In this context, the agent's CPT utility is:

$$\text{CPT}(\mathbf{q}) = \sum_{k=1}^{n} \pi_k^+ V(R(\mathbf{q}, \theta_k))$$

Consider the Lagrangian in (14). Given that $\mathbf{q}^* = (q_1^*, \ldots, q_n^*)$, τ^*, and $\boldsymbol{\mu}^* = (0, \ldots, 0)$ are optimal points, the KKT stationarity conditions are:

$$\frac{\partial L(\mathbf{q}^*, \tau^*, \boldsymbol{\mu}^*)}{\partial q_k^*} = 0, \qquad \text{for } k \in \{1, \ldots, n\}$$

For illustration's sake, consider the partial derivative of L with respect to q_n^*:

$$\frac{\partial L(\mathbf{q}^*, \tau^*, \boldsymbol{\mu}^*)}{\partial q_n^*} = \left(\sum_{k=1}^{n} \pi_k^+ \frac{\partial V(R(\mathbf{q}^*, \theta_k))}{\partial q_n^*} \right) - \tau^* = 0$$

By considering all partial derivatives $\frac{\partial L(\mathbf{q}^*, \tau^*, \boldsymbol{\mu}^*)}{\partial q_k^*}$, for $k \in \{1, \ldots, n\}$, we end up with the following system of equations:

$$\begin{array}{ccccccccc}
\pi_1^+ \frac{\partial V(R(\mathbf{q}^*, \theta_1))}{\partial q_1^*} & + & \pi_2^+ \frac{\partial V(R(\mathbf{q}^*, \theta_2))}{\partial q_1^*} & + \cdots + & \pi_n^+ \frac{\partial V(R(\mathbf{q}^*, \theta_n))}{\partial q_1^*} & = & \tau^* \\
\pi_1^+ \frac{\partial V(R(\mathbf{q}^*, \theta_1))}{\partial q_2^*} & + & \pi_2^+ \frac{\partial V(R(\mathbf{q}^*, \theta_2))}{\partial q_2^*} & + \cdots + & \pi_n^+ \frac{\partial V(R(\mathbf{q}^*, \theta_n))}{\partial q_2^*} & = & \tau^* \\
\vdots & & \vdots & & \vdots & & \vdots \\
\pi_1^+ \frac{\partial V(R(\mathbf{q}^*, \theta_1))}{\partial q_n^*} & + & \pi_2^+ \frac{\partial V(R(\mathbf{q}^*, \theta_2))}{\partial q_n^*} & + \cdots + & \pi_n^+ \frac{\partial V(R(\mathbf{q}^*, \theta_n))}{\partial q_n^*} & = & \tau^*
\end{array} \tag{21}$$

Given the definition of π_k^+ in (7), the system of equations in (21) can then be written as:

$$\begin{bmatrix} -1 & a_{1,2} & \cdots & a_{1,n} \\ -1 & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ -1 & a_{n,2} & \cdots & a_{n,n} \end{bmatrix} \begin{bmatrix} \tau^* \\ W^+\left(\sum_{k=2}^{n} p_k\right) \\ \vdots \\ W^+(p_{n-1} + p_n) \\ W^+(p_n) \end{bmatrix} = \begin{bmatrix} -\frac{\partial V(R(\mathbf{q}, \theta_1))}{\partial q_1^*} \\ -\frac{\partial V(R(\mathbf{q}, \theta_1))}{\partial q_2^*} \\ \vdots \\ -\frac{\partial V(R(\mathbf{q}, \theta_1))}{\partial q_n^*} \end{bmatrix} \tag{22}$$

where $a_{x,k} = \frac{\partial V(R(\mathbf{q},\theta_k))}{\partial q_x^*} - \frac{\partial V(R(\mathbf{q},\theta_{k-1}))}{\partial q_x^*}$. Hence, the system of equations in (21) becomes a system of linear equations with n equations and n unknowns when one considers all the $W^+(\cdot)$ as well as τ^* as variables. If the leftmost matrix is full rank, then this system admits a single solution. We have shown in Corollary 1 that comonotonicity guarantees that any reported belief is associated with a single true belief. This ensures that the above system has a single solution since the relevant variables are the subjective probabilities.

Let $(y_1, \ldots, y_n)$ be the solution to the system of linear equations in (22), where $W^+(p_n) = y_n$. Recall that W^+ is strictly increasing, hence it admits an inverse function. Consequently, $p_n = W^{+^{-1}}(y_n)$. Similarly, let $W^+(p_n + p_{n-1}) = y_{n-1} \implies p_{n-1} = W^{+^{-1}}(y_{n-1}) - p_n = W^{+^{-1}}\left(y_{n-1}\right) - W^{+^{-1}}\left(y_n\right)$. More generally, for all $k \in \{2, \ldots, n-1\}$, we obtain p_k using backward substitution, i.e., by solving the equation $p_k = W^{+^{-1}}\left(y_k\right) - W^{+^{-1}}\left(y_{k+1}\right)$. Lastly, $p_1 = 1 - \sum_{k=2}^{n} p_k$ and, thus, we obtain an agent's true belief.

6.1.1 Numerical example

Consider three exhaustive and mutually exclusive outcomes ($n = 3$) and let $R(\mathbf{q}, \theta_x)$ be the following positive, comonotonic proper scoring rule based on the quadratic scoring rule:

$$
\begin{aligned}
R(\mathbf{q}, \theta_1) &= 2q_1 - q_1^2 - q_2^2 - q_3^2 + 1 \\
R(\mathbf{q}, \theta_2) &= 2q_2 - q_1^2 - q_2^2 - q_3^2 + 3.1 \\
R(\mathbf{q}, \theta_3) &= 2q_3 - q_1^2 - q_2^2 - q_3^2 + 5.2
\end{aligned}
$$

where $R(\mathbf{q}, \theta_3) > R(\mathbf{q}, \theta_2) > R(\mathbf{q}, \theta_1) \geq 0$. Suppose that an agent reports the belief $\mathbf{q}^* = (0.1, 0.3, 0.6)$. For illustration's sake, consider the last equation in (21):

$$
\begin{aligned}
&\left[1 - W^+\left(p_2 + p_3\right)\right] \frac{\partial V(R(\mathbf{q}, \theta_1))}{\partial q_3^*} + \left[W^+(p_2 + p_3) - W^+(p_3)\right] \frac{\partial V(R(\mathbf{q}, \theta_2))}{\partial q_3^*} \\
&\quad + W^+(p_3) \frac{\partial V(R(\mathbf{q}, \theta_3))}{\partial q_3^*} \\
&= \tau^*
\end{aligned}
$$

Let V be the value function in (6) with parameter $\alpha = 0.88$. Consequently, the above equation can be written as follows:

$$
1.4855 \cdot W^+(p_3) + 0.1778 \cdot W^+(p_2 + p_3) - \tau^* = 1.0949
$$

Repeating the same procedure for all partial derivatives $\frac{\partial L(\mathbf{q}^*, \tau^*, \boldsymbol{\mu}^*)}{\partial q_k^*}$, for $k \in \{1, 2, 3\}$, we end up with the following system of linear equations:

$$0.0107 \cdot W^{+}(p_3) - 1.7951 \cdot W^{+}(p_2 + p_3) - \tau^* = -1.6423$$
$$-1.4963 \cdot W^{+}(p_3) + 1.6173 \cdot W^{+}(p_2 + p_3) - \tau^* = 0.5474$$
$$1.4855 \cdot W^{+}(p_3) + 0.1778 \cdot W^{+}(p_2 + p_3) - \tau^* = 1.0949$$

The above system of linear equations is equivalent to the system in (22). After solving the above system of equations, we obtain: $W^{+}(p_3) \approx 0.6271$ and $W^{+}(p_2 + p_3) \approx 0.9186$. Let W^{+} be the weighting function proposed by Tversky and Kahneman (1992) shown in (9) with parameter $\gamma = 0.61$. We then obtain $p_3 = W^{+^{-1}}(0.6271) \approx 0.8224$, $p_2 = W^{+^{-1}}(0.9186) - p_3 \approx 0.169$, and $p_1 = 1 - p_3 - p_2 = 0.0086$. Thus, the agent's true belief is equal to $\mathbf{p} = (0.0086, 0.169, 0.8224)$.

6.1.2 Understanding how value functions and weighting functions distort reported beliefs

To build intuition and to better understand how value functions and weighting functions might distort reported beliefs, consider the simple case with two outcomes ($n = 2$), where the proper scoring rule is the comonotonic scoring rule in (18):

$$R(\mathbf{q}, \theta_1) = 2q_1 - q_1^2 - q_2^2 + 1$$
$$R(\mathbf{q}, \theta_2) = 2q_2 - q_1^2 - q_2^2 + 3.1$$

Recall that $R(\mathbf{q}, \theta_2) > R(\mathbf{q}, \theta_1) \geq 0$. Assume agents have the value function defined in (6). The system of equations in (21) then implies that:

$$\left(1 - W^{+}(p_2)\right) \cdot \alpha \cdot \left(2 - 2q_1\right) \cdot R(\mathbf{q}, \theta_1)^{\alpha-1} + W^{+}(p_2) \cdot \alpha \cdot \left(-2q_1\right) \cdot R(\mathbf{q}, \theta_2)^{\alpha-1} = \tau^*$$
$$\left(1 - W^{+}(p_2)\right) \cdot \alpha \cdot \left(-2q_2\right) \cdot R(\mathbf{q}, \theta_1)^{\alpha-1} + W^{+}(p_2) \cdot \alpha \cdot \left(2 - 2q_2\right) \cdot R(\mathbf{q}, \theta_2)^{\alpha-1} = \tau^*$$

Combining the above equations, we obtain that:

$$\frac{W^{+}(p_2)}{1 - W^{+}(p_2)} = \frac{q_2}{1 - q_2} \cdot \left(\frac{R(\mathbf{q}, \theta_1)}{R(\mathbf{q}, \theta_2)}\right)^{\alpha-1}$$

Since $R(\mathbf{q}, \theta_2) > R(\mathbf{q}, \theta_1) \geq 0$, the fraction $\frac{R(\mathbf{q}, \theta_1)}{R(\mathbf{q}, \theta_2)}$ is positive and always less than or equal to 1. Let $c = \frac{R(\mathbf{q}, \theta_1)}{R(\mathbf{q}, \theta_2)}$. Solving the above equation for q_2, we obtain that:

$$q_2 = \frac{W^{+}(p_2)}{W^{+}(p_2) + \left(1 - W^{+}(p_2)\right) \cdot c^{\alpha-1}}$$

When $\alpha = 1$, i.e., when the utility function is linear, the reported belief is then $q_2 = W^{+}(p_2)$. That is, an agent reports his subjective probability distorted by the weighting function W^{+}. When $\alpha < 1$, which implies risk aversion under the expected utility theory framework, the agent reports $q_2 < W^{+}(p_2)$. Intuitively, when increasing risk aversion, the difference between the rewards in the two underlying outcomes becomes more important, which leads the agent to report less extreme probability values, i.e., the agent reports more uniform probabilities to reduce the variability of his payoff. Finally, when $\alpha > 1$, we have the opposite effect, i.e., the agent reports a more extreme (closer to 1) probability value $q_2 > W^{+}(p_2)$. Note that in all cases, the value function must be known a priori so that $c^{\alpha-1}$ becomes a numerical value. Moreover, the inverse of the weighting function must be known so as to obtain the agent's true belief. For example, when $\alpha = 1$, $p_2 = W^{+^{-1}}(q_2)$. Since weighting functions are continuous and strictly increasing, they always admit inverse functions which are also continuous and strictly increasing. In the following subsections, we generalize the above case by considering any finite number of outcomes and negative and mixed comonotonic proper scoring rules.

6.2 Negative and comonotonic proper scoring rules

Now, suppose that an agent's belief is elicited through a negative, comonotonic proper scoring rule R, for $0 \geq R(\mathbf{q}, \theta_n) > R(\mathbf{q}, \theta_{n-1}) > \cdots > R(\mathbf{q}, \theta_2) > R(\mathbf{q}, \theta_1)$. The CPT utility of that agent is:

$$\mathrm{CPT}(\mathbf{q}) = \sum_{k=1}^{n} \pi_k^- V(R(\mathbf{q}, \theta_k))$$

By considering all partial derivatives $\frac{\partial L(\mathbf{q}^*, \tau^*, \boldsymbol{\mu}^*)}{\partial q_k^*}$, for $k \in \{1, \ldots, n\}$, we end up with the following system of equations:

$$\begin{aligned}
\pi_1^- \frac{\partial V(R(\mathbf{q}^*, \theta_1))}{\partial q_1^*} + \pi_2^- \frac{\partial V(R(\mathbf{q}^*, \theta_2))}{\partial q_1^*} + \cdots + \pi_n^- \frac{\partial V(R(\mathbf{q}^*, \theta_n))}{\partial q_1^*} &= \tau^* \\
\pi_1^- \frac{\partial V(R(\mathbf{q}^*, \theta_1))}{\partial q_2^*} + \pi_2^- \frac{\partial V(R(\mathbf{q}^*, \theta_2))}{\partial q_2^*} + \cdots + \pi_n^- \frac{\partial V(R(\mathbf{q}^*, \theta_n))}{\partial q_2^*} &= \tau^* \\
\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad & \vdots \\
\pi_1^- \frac{\partial V(R(\mathbf{q}^*, \theta_1))}{\partial q_n^*} + \pi_2^- \frac{\partial V(R(\mathbf{q}^*, \theta_2))}{\partial q_n^*} + \cdots + \pi_n^- \frac{\partial V(R(\mathbf{q}^*, \theta_n))}{\partial q_n^*} &= \tau^*
\end{aligned} \tag{23}$$

Given the definition of π_k^- in (8), the system of equations in (23) can then be written as:

$$\begin{bmatrix} b_{1,1} & \cdots & b_{1,n-1} & -1 \\ b_{2,1} & \cdots & b_{2,n-1} & -1 \\ \vdots & \ddots & \vdots & \vdots \\ b_{n,1} & \cdots & b_{n,n-1} & -1 \end{bmatrix} \begin{bmatrix} W^{-}(p_1) \\ W^{-}(p_1+p_2) \\ \vdots \\ W^{-}\left(\sum_{k=1}^{n-1} p_k\right) \\ \tau^* \end{bmatrix} = \begin{bmatrix} -\frac{\partial V(R(\mathbf{q},\theta_n))}{\partial q_1^*} \\ -\frac{\partial V(R(\mathbf{q},\theta_n))}{\partial q_2^*} \\ \vdots \\ -\frac{\partial V(R(\mathbf{q},\theta_n))}{\partial q_n^*} \end{bmatrix} \tag{24}$$

where $b_{x,k} = \frac{\partial V(R(\mathbf{q},\theta_k))}{\partial q_x^*} - \frac{\partial V(R(\mathbf{q},\theta_{k+1}))}{\partial q_x^*}$. Hence, the system of equations in (23) becomes a system of linear equations with n equations and n unknowns when one considers all the the $W^{-}(\cdot)$ as well as τ^* as variables. Let $(y_1, \dots, y_n)$ be the solution to such a system of linear equations, where $W^{-}(p_1) = y_1$. The uniqueness of such a solution is guaranteed by the comonotonicity property (see Corollary 2). Since W^{-} is strictly increasing, it admits an inverse function. Consequently, $p_1 = W^{-^{-1}}(y_1)$. Similarly, let $W^{-}(p_1 + p_2) = y_2 \implies p_2 = W^{-^{-1}}(y_2) - p_1 = W^{-^{-1}}(y_2) - W^{-^{-1}}(y_1)$. More generally, for all $k \in \{2, \dots, n-1\}$, we obtain p_k using forward substitution, i.e., by solving the equation $p_k = W^{-^{-1}}(y_k) - W^{-^{-1}}(y_{k-1})$. Lastly, $p_n = 1 - \sum_{k=1}^{n-1} p_k$ and, thus, we obtain a CPT agent's true belief.

6.2.1 Numerical example

Consider three exhaustive and mutually exclusive outcomes ($n = 3$) and and let $R(\mathbf{q}, \theta_x)$ be the following negative, comonotonic proper scoring rule based on the quadratic scoring rule:

$$R(\mathbf{q}, \theta_1) = 2q_1 - q_1^2 - q_2^2 - q_3^2 - 5.2$$
$$R(\mathbf{q}, \theta_2) = 2q_2 - q_1^2 - q_2^2 - q_3^2 - 3.1$$
$$R(\mathbf{q}, \theta_3) = 2q_3 - q_1^2 - q_2^2 - q_3^2 - 1$$

where $0 \geq R(\mathbf{q}, \theta_3) > R(\mathbf{q}, \theta_2) > R(\mathbf{q}, \theta_1)$. Suppose that an agent reports the belief $\mathbf{q}^* = (0.1, 0.3, 0.6)$. For illustration's sake, consider the last equation in (23):

$$\begin{aligned} W^{-}(p_1)\frac{\partial V(R(\mathbf{q}, \theta_1))}{\partial q_3^*} &+ \left[W^{-}(p_1+p_2) - W^{-}(p_1)\right]\frac{\partial V(R(\mathbf{q}, \theta_2))}{\partial q_3^*} \\ &+ \left[1 - W^{-}\left(p_1+p_2\right)\right]\frac{\partial V(R(\mathbf{q}, \theta_3))}{\partial q_3^*} \\ = \tau^* \end{aligned}$$

Let V be equal to the value function in (6) with parameters $\beta = 0.88$ and $\lambda = 2.25$. Consequently, the above equation can be written as follows:

$$0.1478 \cdot W^{-}(p_1) - 3.9478 \cdot W^{-}(p_1 + p_2) - \tau^* = -1.8619$$

Repeating the same procedure for all partial derivatives $\frac{\partial L(\mathbf{q}^*,\tau^*,\boldsymbol{\mu}^*)}{\partial q_k^*}$, for $k \in \{1,2,3\}$, we end up with the following system of linear equations:

$$\begin{aligned}
3.2548 \cdot W^-(p_1) + 0.1178 \cdot W^-(p_1+p_2) - \tau^* &= 0.4655 \\
-3.4026 \cdot W^-(p_1) + 3.83 \cdot W^-(p_1+p_2) - \tau^* &= 1.3964 \\
0.1478 \cdot W^-(p_1) - 3.9478 \cdot W^-(p_1+p_2) - \tau^* &= -1.8619
\end{aligned}$$

The above system of linear equations is equivalent to the system in (24). After solving the above system of equations, we obtain: $W^-(p_1) \approx 0.1258$ and $W^-(p_1+p_2) \approx 0.4763$. Let W^- be the weighting function proposed by Tversky and Kahneman (1992) shown in (9) with parameter $\delta = 0.69$. We then obtain $p_1 = W^{-^{-1}}(0.1258) \approx 0.0608$, $p_2 = W^{-^{-1}}(0.4763) - p_1 \approx 0.4746$, and $p_3 = 1 - p_1 - p_2 = 0.4646$. Thus, the agent's true belief is equal to $\mathbf{p} = (0.0608, 0.4746, 0.4646)$.

6.3 Mixed and comonotonic proper scoring rules

Finally, suppose that an agent's belief is elicited through a mixed, comonotonic proper scoring rule R, for $R(\mathbf{q},\theta_n) > R(\mathbf{q},\theta_{n-1}) > \cdots > R(\mathbf{q},)$. The CPT utility of that agent is then:

$$\mathrm{CPT}(\mathbf{q}) = \sum_{k=1}^{i-1} \pi_k^- V(R(\mathbf{q},\theta_k)) + \sum_{k=i}^{n} \pi_k^+ V(R(\mathbf{q},\theta_k))$$

By considering all partial derivatives $\frac{\partial L(\mathbf{q}^*,\tau^*,\boldsymbol{\mu}^*)}{\partial q_k^*}$, for $k \in \{1,\dots,n\}$, we end up with the following system of equations:

$$\begin{aligned}
\sum_{k=1}^{i-1} \pi_k^- \frac{\partial V(R(\mathbf{q},\theta_k))}{\partial q_1^*} + \sum_{k=i}^{n} \pi_k^+ \frac{\partial V(R(\mathbf{q},\theta_k))}{\partial q_1^*} &= \tau^* \\
\sum_{k=1}^{i-1} \pi_k^- \frac{\partial V(R(\mathbf{q},\theta_k))}{\partial q_2^*} + \sum_{k=i}^{n} \pi_k^+ \frac{\partial V(R(\mathbf{q},\theta_k))}{\partial q_2^*} &= \tau^* \\
\vdots \qquad\qquad\qquad \vdots \qquad &\quad \vdots \\
\sum_{k=1}^{i-1} \pi_k^- \frac{\partial V(R(\mathbf{q},\theta_k))}{\partial q_n^*} + \sum_{k=i}^{n} \pi_k^+ \frac{\partial V(R(\mathbf{q},\theta_k))}{\partial q_n^*} &= \tau^*
\end{aligned} \tag{25}$$

For a mixed proper scoring rule, the system of equations in (25) can be written as:

$$\begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,i-1} & a_{1,i} & \cdots & a_{1,n} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,i-1} & a_{2,i} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{n,1} & b_{n,2} & \cdots & b_{n,i-1} & a_{n,i} & \cdots & a_{n,n} \end{bmatrix} \begin{bmatrix} W^-(p_1) \\ W^-(p_1 + p_2) \\ \vdots \\ W^-\left(\sum_{k=1}^{i-1} p_k\right) \\ W^+\left(\sum_{k=i}^{n} p_k\right) \\ \vdots \\ W^+\left(p_n + p_{n-1}\right) \\ W^+\left(p_n\right) \end{bmatrix} = \begin{bmatrix} \tau^* \\ \tau^* \\ \vdots \\ \tau^* \end{bmatrix} \tag{26}$$

where $a_{x,k} = \frac{\partial V(R(\mathbf{q},\theta_k))}{\partial q_x^*} - \frac{\partial V(R(\mathbf{q},\theta_{k-1}))}{\partial q_x^*}$, and $b_{x,k} = \frac{\partial V(R(\mathbf{q},\theta_k))}{\partial q_x^*} - \frac{\partial V(R(\mathbf{q},\theta_{k+1}))}{\partial q_x^*}$. By considering all the $W(\cdot)$ as well as τ^* as variables, the above system of equations becomes a system of linear equations with $n+1$ variables and n equations. Consequently, a priori, this system has no solution. In what follows, we describe a procedure to obtain the value of τ^* and, thus, to reduce the number of unknowns by one. The procedure consists of three steps:

(1) Solve (26) for $W^-\left(\sum_{k=1}^{i-1} p_k\right)$ in terms of τ^*, i.e., $W^-\left(\sum_{k=1}^{i-1} p_k\right) = y_{i-1} \cdot \tau^*$ $\implies \sum_{k=1}^{i-1} p_k = W^{-^{-1}}(y_{i-1} \cdot \tau^*)$, where $y_{i-1} \in \mathfrak{R}^+$ is a known numerical value;

(2) Solve (26) for $W^+\left(\sum_{k=i}^{n} p_k\right)$ in terms of τ^*, i.e., $W^+(\sum_{k=i}^{n} p_k) = y_i \cdot \tau^*$ $\implies \sum_{k=i}^{n} p_k = W^{+^{-1}}(y_i \cdot \tau^*)$, where $y_i \in \mathfrak{R}^+$ is a known numerical value;

(3) Combine the results from the first and second steps to obtain τ^*, i.e., $\sum_{k=1}^{i-1} p_k + \sum_{k=i}^{n} p_k = W^{-^{-1}}(y_{i-1} \cdot \tau^*) + W^{+^{-1}}(y_i \cdot \tau^*) \implies 1 = W^{-^{-1}}(y_{i-1} \cdot \tau^*)$ $+W^{+^{-1}}(y_i \cdot \tau^*)$.

Given that W^+, W^-, y_i, and y_{i-1} are all known, the value of τ^* can be computed numerically after the third step. We discuss an approach to do so in the numerical example below. The value of τ^* always exists due to the Intermediate Value Theorem. In particular, let $H(\tau^*) = W^{-1}(y_{i-1} \cdot \tau^*) + W^{-1}(y_{i-1} \cdot \tau^*)$. Note that H is continuous, $H(0) = 0$, and $H\left(\min\left(\frac{1}{y_{i-1}}, \frac{1}{y_i}\right)\right) > 1$. Hence, $H(0) < 1 < H(\min(\frac{1}{y_{i-1}}, \frac{1}{y_i}))$, which according to the Intermediate Value Theorem implies that there exists a τ^* such that $H(\tau^*) = 1$. After finding τ^*, the system of equations in (26) becomes a system of linear equations with n variables and n equations. The true belief $\mathbf{p}$ can then be obtained by following the four extra steps described below:

(4) Let $W^-(p_1) = y_1 \cdot \tau^*$. Consequently, $p_1 = W^{-^{-1}}(y_1 \cdot \tau^*)$;

(5) Let $W^+(p_n) = y_n \cdot \tau^*$. Consequently, $p_n = W^{+^{-1}}(y_n \cdot \tau^*)$;

(6) For all $k \in \{2, \ldots, i-1\}$, we obtain p_k as described in Sect. 6.2, i.e., by solving the equation $p_k = W^{-^{-1}}\left(y_k \cdot \tau^*\right) - W^{-^{-1}}\left(y_{k-1} \cdot \tau^*\right)$;

(7) For all $k \in \{i, \ldots, n-1\}$, we obtain p_k as described in Sect. 6.1, i.e., by solving the equation $p_k = W^{+^{-1}}\left(y_k \cdot \tau^*\right) - W^{+^{-1}}\left(y_{k+1} \cdot \tau^*\right)$.

The above solution relies on the resulting leftmost matrix in (26) being full rank. In other words, it relies on each reported belief being associated with a single true belief, which is guaranteed by the comonotonicity property we proved in Proposition 1.

6.3.1 Numerical example

Consider three exhaustive and mutually exclusive outcomes ($n = 3$) and let $R(\mathbf{q}, \theta_x)$ be the following mixed, comonotonic proper scoring rule based on the quadratic scoring rule:

$$\begin{aligned} R(\mathbf{q}, \theta_1) &= 2q_1 - q_1^2 - q_2^2 - q_3^2 - 1 \\ R(\mathbf{q}, \theta_2) &= 2q_2 - q_1^2 - q_2^2 - q_3^2 + 3.1 \\ R(\mathbf{q}, \theta_3) &= 2q_3 - q_1^2 - q_2^2 - q_3^2 + 5.2 \end{aligned}$$

where $R(\mathbf{q}, \theta_3) > R(\mathbf{q}, \theta_2) > 0 \geq R(\mathbf{q}, \theta_1)$. Suppose that an agent reports the belief $\mathbf{q}^* = (0.1, 0.3, 0.6)$. For illustration's sake, consider the last equation in (25):

$$\begin{aligned} & W^-(p_1)\frac{\partial V(R(\mathbf{q}, \theta_1))}{\partial q_3^*} + \left[W^+\left(p_2 + p_3\right) - W^+\left(p_3\right)\right]\frac{\partial V(R(\mathbf{q}, \theta_2))}{\partial q_3^*} \\ & \quad + W^+(p_3)\frac{\partial V(R(\mathbf{q}, \theta_3))}{\partial q_3^*} \\ & = \tau^* \end{aligned}$$

Let V be equal to the value function in (6) with parameters $\alpha = \beta = 0.88$ and $\lambda = 2.25$. Consequently, the above equation can be written as follows:

$$-2.311 \cdot W^-(p_1) - 0.9171 \cdot W^+(p_2 + p_3) + 1.4855 \cdot W^+(p_3) = \tau^*$$

Repeating the same procedure for all partial derivatives $\frac{\partial L(\mathbf{q}^*, \tau^*, \mu^*)}{\partial q_k^*}$, for $k \in \{1, 2, 3\}$, we end up with the following system of linear equations:

$$\begin{aligned} & 3.4665 \cdot W^-(p_1) - 0.1528 \cdot W^+(p_2 + p_3) + 0.0107 \cdot W^+(p_3) = \tau^* \\ & -1.1555 \cdot W^-(p_1) + 1.0699 \cdot W^+(p_2 + p_3) - 1.4963 \cdot W^+(p_3) = \tau^* \\ & -2.311 \cdot W^-(p_1) - 0.9171 \cdot W^+(p_2 + p_3) + 1.4855 \cdot W^+(p_3) = \tau^* \end{aligned}$$

The above system of linear equations is equivalent to the system in (26) for $i = 2$. After solving the above system of equations, we obtain: $W^-(p_1) = -1844.259 \cdot \tau^*$, $W^+(p_2 + p_3) = -43947.14 \cdot \tau^*$, and $W^+(p_3) = -30000 \cdot \tau^*$. Thus, $p_1 = W^{-^{-1}}(-1844.259 \cdot \tau^*)$, $p_2 + p_3 = W^{+^{-1}}(-43947.14 \cdot \tau^*)$, and then $1 = W^{-^{-1}}(-1844.259 \cdot \tau^*) + W^{+^{-1}}(-43947.14 \cdot \tau^*)$.

For ease of exposition, let $\psi = -1844.259 \cdot \tau^*$, and $\phi = -43947.14 \cdot \tau^*$. By construction, $\psi, \phi \in [0, 1]$. Moreover, $\phi = \frac{43947.14}{1844.259} \cdot \psi$. Thus, to find the value of τ^*, we just need to find the value of $\psi \in [0, 1]$ such that $1 = W^{-^{-1}}(\psi) + W^{+^{-1}}\left(\frac{43947.14}{1844.259} \cdot \psi\right)$. Let W^+ and W^- be the weighting functions proposed by Tversky and Kahneman (1992) shown in (9) with parameters $\gamma = 0.61$ and $\delta = 0.69$. We find numerically that $\psi \approx 0.0384$ and, consequently, $\tau^* = -0.00002082$. Finally, we obtain that $p_1 = W^{-^{-1}}(-1844.259 \cdot \tau^*) \approx 0.0095$, $p_3 = W^{+^{-1}}(-30000 \cdot \tau^*) \approx 0.8197$, and $p_2 = W^{+^{-1}}(-43947.14 \cdot \tau^*) - p_3 \approx 0.1708$. Thus, the agent's true belief is equal to $\mathbf{p} = (0.0095, 0.1708, 0.8197)$.

6.4 A comment on the numerical examples in Sects. 6.1, 6.2, and 6.3

Table 1 summarizes the results from our numerical examples in Sects. 6.1, 6.2, and 6.3. In particular, it highlights how cumulative prospect theory agents with different beliefs end up reporting the same belief under different comonotonic proper scoring rules. The same would not be true if agents behaved to maximize their expected scores. This highlights the risk of assuming a wrong decision model when eliciting beliefs.

7 Discussion

Proper scoring rules have been widely used to induce honest reporting of subjective probabilities. The main assumption behind proper scoring rules is that agents are risk neutral, which is often an unrealistic assumption when the underlying agents are humans. In this paper, we adapted proper scoring rules to cumulative prospect theory (Tversky and Kahneman 1992), a more modern model of decision making

Table 1 True beliefs of cumulative prospect theory agents reporting (0.1, 0.3, 0.6) under positive, negative, and mixed comonotonic proper scoring rules

	Comonotonic proper scoring rule	True belief (p)
Positive	$R(\mathbf{q}, \theta_1) = 2q_1 - q_1^2 - q_2^2 - q_3^2 + 1$	(0.0086, 0.169, 0.8224)
	$R(\mathbf{q}, \theta_2) = 2q_2 - q_1^2 - q_2^2 - q_3^2 + 3.1$	
	$R(\mathbf{q}, \theta_3) = 2q_3 - q_1^2 - q_2^2 - q_3^2 + 5.2$	
Negative	$R(\mathbf{q}, \theta_1) = 2q_1 - q_1^2 - q_2^2 - q_3^2 - 5.2$	(0.0608, 0.4746, 0.4646)
	$R(\mathbf{q}, \theta_2) = 2q_2 - q_1^2 - q_2^2 - q_3^2 - 3.1$	
	$R(\mathbf{q}, \theta_3) = 2q_3 - q_1^2 - q_2^2 - q_3^2 - 1$	
Mixed	$R(\mathbf{q}, \theta_1) = 2q_1 - q_1^2 - q_2^2 - q_3^2 - 1$	(0.0095, 0.1708, 0.8197)
	$R(\mathbf{q}, \theta_2) = 2q_2 - q_1^2 - q_2^2 - q_3^2 + 3.1$	
	$R(\mathbf{q}, \theta_3) = 2q_3 - q_1^2 - q_2^2 - q_3^2 + 5.2$	

under uncertainty. In particular, we started by showing how traditional proper scoring rules are naturally incompatible with cumulative prospect theory. This happens because multiple true beliefs might be associated with a single reported belief when agents report under traditional proper scoring rules and behave according to cumulative prospect theory. This implies that the reported belief cannot always be used to uniquely determine the correct true belief. We also explain that this problem does not only happen around the probability value of 0.5. As illustrated in the paper, this problem is dependent on the underlying proper scoring rule.

Next, we proved that comonotonicity is a sufficient condition to fix the above problem. We covered all possible scenarios regarding the underlying comonotonic proper scoring rule, i.e., when the proper scoring rule is positive, negative, and mixed. Moreover, our discussion is valid for any finite number of outcomes. We also suggested how to construct a comonotonic proper scoring rule from any bounded proper scoring rule. When combined, our results generalize previous work by Offerman et al. (2009) and Kothiyal et al. (2011) when it comes to the use of proper scoring rules for eliciting beliefs of CPT agents.

Finally, we proposed procedures to obtain a CPT agent's true belief from his reported belief, when beliefs are elicited by means of comonotonic proper scoring rules. Our suggested procedures involve solving systems of linear equations when all the components that drive the agent's attitude towards uncertainty are known. Hence, an assumption we made was that the components that drive an agent's attitude towards uncertainty are known a priori, i.e., one must elicit an agent's value function and weighting functions before eliciting the agent's subjective probabilities. A question that arises is: is the elicitation of value/weighting functions really necessary? If one wants to make deterministic payments, then the answer to this question is "yes". In particular, Schlag and van der Weele (2013) proved that if agents are not risk neutral, then it is not possible to elicit subjective probabilities or the mean of a subjective probability distribution truthfully using deterministic payment schemes. Clearly, after eliciting value/weighting functions, the requester interested in the elicitation of beliefs could very well incorporate the obtained value/weighting functions into a scoring rule so as to tailor the same to a certain agent's risk attitudes. In spirit, this is what Winkler and Murphy (1970) suggested for expected utility maximizers. In that case, no calibration of reported beliefs is required a posteriori since the tailored proper scoring rule will already induce an agent to report his true belief. In this paper, we decided to follow the calibration approach because it better highlights the importance of comonotonicity when eliciting beliefs from CPT agents.

The results from the experiments by Armantier and Treich (2013) demonstrate why a mechanism such as the one we propose in this paper is desirable when eliciting beliefs using proper scoring rules. Specifically, using a positive variant of the quadratic scoring rule, Armantier and Treich (2013) empirically found that risk aversion leads agents to report more uniform probabilities. This is a rather safe strategy because the reporting agent will always receive some monetary payment no matter which outcome occurs. In this case, the marginal utility of the monetary payment to the agent confounds the effect of his belief. This shows why it is important to know and remove the influence of all the components that drive risk attitudes on agents' reported beliefs, which is precisely what our proposed mechanism does.

An alternative to using deterministic payments is to make payments in lottery tickets. There have been some suggestions on how probabilistic payments based on proper scoring rules can elicit an agent's belief when the components that drive the agent's attitude towards uncertainty are unknown (Allen 1987; Karni 2009; Hossain and Okui 2013; Schlag and van der Weele 2013; Sandroni and Shmaya 2013; Carvalho 2016a). There are, however, doubts as to the effectiveness of this approach since it relies on the assumption of expected value maximization in tickets. Selten et al. (1999) showed that significantly more violations of expected value maximization occur when payments were issued in lottery tickets as opposed to direct money payoffs. The authors further observed that anomalies such as the common ratio effect, the preference reversal effect, and violations of stochastic dominance are further exacerbated when agents receive payments in lottery tickets.

The results by Hossain and Okui (2013) differ from the results by Selten et al. (1999) in that, on the aggregate level, their lottery-based payment scheme induced agents to report beliefs closer to their true beliefs than the quadratic scoring rule. We note, however, that the effectiveness of the approach by Hossain and Okui (2013) in inducing honest reporting is, to a certain degree, dependent on agents' risk attitudes. For example, the authors found no significant difference between their approach and the quadratic scoring rule when agents are risk neutral and, surprisingly, the quadratic scoring rule outperformed their proposed method when agents are risk seeking [see the last column of Table 4 in the paper by Hossain and Okui (2013)]. All in all, it is still unclear whether probabilistic payments are actually effective in inducing honest reporting.

Another alternative to obtain an off-the-shelf method to elicit beliefs is to assume certain functionals, such as Tversky and Kahneman's value and weighting functions, and to instantiate their parameters based on averages of previous empirical findings. Offerman and Palley (2016) obtained remarkably accurate results (on the aggregate level) by following this approach, in a sense that average calibrated beliefs closely matched objective probabilities. We believe that when aggregate beliefs are of more interest than single beliefs, then a similar approach can be successfully used in conjunction with our proposed methods as well, thus eliminating the need of eliciting individual weighting functions and value functions.

An interesting open question involves investigating the most appropriate decision model one should assume when eliciting beliefs using proper scoring rules since the implications of choosing an inappropriate model can be rather drastic. For example, Chambers (2008) showed that when an agent who behaves according to the max–min expected utility model is confronted with a proper scoring rule, that agent will report a belief corresponding to some probability measure in his set of priors. Furthermore, for any prior in that set, there is a different scoring rule that induces the agent to announce such a prior. One can interpret this result as proper scoring rules not being an effective technique to elicit beliefs from max-min expected utility agents.

Despite some well-known problems, such as not accounting for coalescing and violations of stochastic dominance (Birnbaum 2008), cumulative prospect theory has been very successful at describing how humans make decisions under uncertainty (Camerer 2004). Although reporting beliefs under proper scoring rules is theoretically equivalent to making decisions under uncertainty, we note, however, that previous experimental results from decision theory/analysis do not necessarily translate into equivalent results in this elicitation setting. The reason for this is that, as opposed to choosing amongst a finite set of fixed prospects, agents have some control over their potential payoffs under proper scoring rules, which are defined by the reported beliefs. This fact might affect how agents reason under uncertainty, and it asks for a new set of experiments so as to determine the most appropriate decision theory one should assume.

Acknowledgements The authors thank Jonathan Baron and two anonymous reviewers for useful comments. The authors also thank the Natural Sciences and Engineering Research Council of Canada for funding this research.

References

Abdellaoui M (2000) Parameter-free elicitation of utility and probability weighting functions. Manage Sci 46(11):1497–1512

Abdellaoui M, Bleichrodt H, L'Haridon O (2008) A tractable method to measure utility and loss aversion under prospect theory. J Risk Uncert 36(3):245–266

Allais M (1953) Violations of the betweenness axiom and nonlinearity in probability. Econometrica 21:503–546

Allen F (1987) Discovering personal probabilities when utility functions are unknown. Manage Sci 33(4):542–544

Armantier O, Treich N (2013) Eliciting beliefs: proper scoring rules, incentives, stakes and hedging. Eur Econ Rev 62:17–40

Barberis N, Xiong W (2009) What drives the disposition effect? An analysis of a long-standing preference-based explanation. J Fin 64(2):751–784

Bickel JE (2010) Scoring rules and decision analysis education. Decis Anal 7(4):346–357

Birnbaum MH (1997) Violations of monotonicity in judgment and decision making. In: Marley AAJ (ed) Choice, decision, and measurement: essays in Honor of R. Duncan Luce, pp 73–100

Birnbaum MH (2008) New paradoxes of risky decision making. Psychol Rev 115(2):463–501

Camerer CF (2004) Prospect theory in the wild: evidence from the field. In: Camerer CF, Loewenstein G, Rabin M (eds) Advances in behavioral economics, chapter 5, pp 148–161

Carvalho A (2015) Tailored proper scoring rules elicit decision weights. Judgm Decis Mak 10(1):86–96

Carvalho A (2016a) A note on Sandroni–Shmaya belief elicitation mechanism. J Predict Mark 10(2):14–21

Carvalho A (2016b) An overview of applications of proper scoring rules. Decis Anal 13(4):223–242

Carvalho A (2017) On a participation structure that ensures representative prices in prediction markets. Decis Support Syst 104:13–25

Carvalho A, Larson K (2012) Sharing rewards among strangers based on peer evaluations. Decis Anal 9(3):253–273

Chambers CP (2008) Proper scoring rules for general decision models. Games Econ Behav 63(1):32–40

Friedman D (1983) Effective scoring rules for probabilistic forecasts. Manage Sci 29(4):447–454

Gilboa I (1987) Expected utility with purely subjective non-additive probabilities. J Math Econ 16(1):65–88

Gneiting T, Raftery AE (2007) Strictly Proper scoring rules, prediction, and estimation. J Am Stat Assoc 102(477):359–378

Gonzalez R, Wu G (1999) On the shape of the probability weighting function. Cogn Psychol 38:129–166

Hanson R (2003) Combinatorial information market design. Inf Syst Front 5(1):107–119

Hines G, Larson K (2010) Preference elicitation for risky prospects. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, pp 889–896

Hossain T, Okui R (2013) The binarized scoring rule. Rev Econ Stud 80(3):984–1001

Johnstone DJ, Jose VRR, Winkler RL (2011) Tailored scoring rules for probabilities. Decis Anal 8(4):256–268

Kadane JB, Winkler RL (1988) Separating probability elicitation from utilities. J Am Stat Assoc 83(402):357–363

Karni E (2009) A mechanism for eliciting probabilities. Econometrica 77(2):603–606

Kothiyal A, Spinu V, Wakker PP (2011) Comonotonic proper scoring rules to measure ambiguity and subjective beliefs. J Multi-Crit Decis Anal 17(3–4):101–113

Nakazono Y (2013) Strategic behavior of federal open market committee board members: evidence from members' forecasts. J Econ Behav Organ 93:62–70

Offerman T, Palley AB (2016) Lossed in translation: an off-the-shelf method to recover probabilistic beliefs from loss-averse agents. Exp Econ 19(1):1–30

Offerman T, Sonnemans J, Van De Kuilen G, Wakker PP (2009) A truth serum for Non-Bayesians: correcting proper scoring rules for risk attitudes. Rev Econ Stud 76(4):1461–1489

Parkes DC, Wellman MP (2015) Economic reasoning and artificial intelligence. Science 349(6245):267–272

Perny P, Viappiani P, Boukhatem A (2016) Incremental preference elicitation for decision making under risk with the rank-dependent utility model. Uncert Artif Intell

Quiggin J (1982) A theory of anticipated utility. J Econ Behav Organ 3(4):323–343

Sandroni A, Shmaya E (2013) Eliciting beliefs by paying in chance. Econ Theory Bull 1(1):33–37

Savage LJ (1971) Elicitation of personal probabilities and expectations. J Am Stat Assoc 66(336):783–801

Schervish MJ (1989) A general method for comparing probability assessors. Ann Stat 17(4):1856–1879

Schlag KH, Tremewan J, Van der Weele JJ (2015) A penny for your thoughts: a survey of methods for eliciting beliefs. Exp Econ 3(1):457–490

Schlag KH, van der Weele JJ (2013) Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality. Theor Econ Lett 3(1):38–42

Schmeidler D (1989) Subjective probability and expected utility without additivity. Econometrica 57(3):571–587

Schotter A, Trevino I (2014) Belief elicitation in the laboratory. Annu Rev Econ 6:103–128

Selten R, Sadrieh A, Abbink K (1999) Money does not induce risk neutral behavior, but binary lotteries do even worse. Theor Decis 46(3):211–249

Spiegelhalter DJ (1986) Probabilistic prediction in patient management and clinical trials. Stat Med 5(5):421–433

Starmer C (2000) Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. J Econ Lit 38(2):332–382

Tetlock PE (2005) Expert political judgment: how good is it? How can we know? Princeton University Press

Toda M (1963) Measurement of subjective probability distributions. Technical Report ESD-TDR-63-407, Decision Sciences Laboratory, Electronic Systems Division, Air Force Systems Command, United States Air Force, Bedford, MA

Tversky A, Kahneman D (1992) Advances in prospect theory: cumulative representation of uncertainty. J Risk Uncert 5(4):297–323

Wakker P, Deneffe D (1996) Eliciting von Neumann–Morgenstern utilities when probabilities are distorted or unknown. Manage Sci 42(8):1131–1150

Wakker PP (2010) Prospect theory: for risk and ambiguity. Cambridge university press

Weber BJ, Chapman GB (2005) Playing for peanuts: why is risk seeking more common for low-stakes gambles? Organ Behav Hum Decis Process 97(1):31–46

Winkler RL, Murphy AH (1968) "Good" probability assessors. J Appl Meteorol 7(5):751–758

Winkler RL, Murphy AH (1970) Nonlinear utility and the probability score. J Appl Meteorol 9:143–148