Theory Dec. (2018) 84:11–27
https://doi.org/10.1007/s11238-017-9607-y

Empirical evaluation of third-generation prospect theory

Michael H. Birnbaum[1]

Published online: 6 June 2017

Abstract Third generation prospect theory (Schmidt et al. J Risk Uncertain 36:203–223, 2008) is a theory of choices and of judgments of highest buying and lowest selling prices of risky prospects, i.e., of willingness to pay (WTP) and willingness to accept (WTA). The gap between WTP and WTA is sometimes called the "endowment effect" and was previously called the "point of view" effect. Third generation prospect theory (TGPT) combines cumulative prospect theory for risky prospects with the theory that judged values are based on the integration of price paid or price received with the consequences of gambles. In TGPT, the discrepancy between WTP and WTA is due to loss aversion-losses have greater absolute utility than gains of the same value. TGPT was developed independently of similar developments by Birnbaum and Zimmermann (Organ Behav Hum Decis Process 74(2):145–187, 1998) and Luce (Utility of gains and losses: measurement-theoretical and experimental approaches. Erlbaum, Mahwah, 2000). This paper reviews theoretical and empirical findings, to show that TGPT fails as a descriptive model of both choices and judgments. Evidence refutes three implications of TGPT, but they are consistent with configural weight models (Birnbaum and Stegner, J Personal Soc Psychol 37:48–74, 1979) in which loss aversion is not needed to describe the results. In the configural weight models, buyers place greater weight on lower consequences, attributes or estimates of value compared to sellers, who place greater configural weight on higher aspects of an object or prospect.

Support was received from National Science Foundation Grant, SES DRMS-0721126. I thank Ulrich Schmidt for comments on an earlier draft.

✉ Michael H. Birnbaum
mbirnbaum@fullerton.edu

[1] Department of Psychology, California State University, Fullerton, CSUF H-830M, P.O. Box 6846, Fullerton, CA 92834-6846, USA

Keywords Cumulative prospect theory · Endowment effect · Loss aversion · Prospect theory · Configural weighting · Willingness to pay (WTP) · Willingness to accept (WTA)

1 Introduction

Schmidt et al. (2008) proposed third-generation prospect theory (TGPT) as a unified theory to account for judgments of value of risky prospects as well as choices between such prospects. This theory was intended to account for the discrepancy between willingness to pay (WTP) and willingness to accept (WTA) and for preference reversals between choices and judgments of value.

This paper shows that third-generation prospect theory implies three properties that are empirically violated by data but which are consistent with an older theory known as configural weight theory. Before presenting the properties and the empirical evidence testing them, it is useful to review the history of the models and their relationships.

Original prospect theory (Kahneman and Tversky 1979) made use of a subjectively weighted utility formulation similar to that of Edwards (1954). Cumulative prospect theory (CPT) by Tversky and Kahneman (1992) is a variant of rank- and sign-dependent utility (RSDU), by Luce and Fishburn (1991), with particular functions specified. Schmidt et al. (2008) refer to CPT as "second generation" prospect theory. Schmidt et al. retained CPT for choices between risky prospects, but added new assumptions to account for judgments of value-maximal buying prices (WTP) and minimal selling prices (WTA). Thus, TGPT is an extension of CPT, rather than a revision of it.

Birnbaum and Stegner (1979) tested a configural weight model that predicted specific relationships between judgments of highest buying and lowest selling prices. Birnbaum and Stegner referred to the empirical effects as effects of the "judge's point of view". They fit a configural weight averaging model in which the configural weights of lower or higher values are affected by instructions to identify with the buyer, seller, or an independent. They assumed that buyers would place greater configural weight on lower estimates, attributes, or consequences of an option than would sellers. The data showed strong effects consistent with this interpretation.

Thaler (1980), who did not cite the earlier configural weight theory or data, proposed the term, "endowment effect" and suggested that such phenomena might relate to "loss aversion," postulated by Kahneman and Tversky (1979) as a property of the utility function—"losses loom larger than gains". The phenomenon is sometimes described as a special case of a "status quo bias" (Samuelson and Zeckhauser 1988). Tversky and Kahneman (1991) elaborated the idea that the discrepancy between WTP and WTA for riskless goods might be explained by a utility function in which a loss of x has greater negative utility than that of a gain of comparable absolute value. Loss aversion was also incorporated in CPT (Tversky and Kahneman 1992).

To represent risky gambles, Schmidt et al. (2008) proposed TGPT, which used CPT, combined with the assumption that prices paid or accepted are integrated into the consequences of a prospect. In TGPT, buying or selling prices of risky prospects are theorized as decisions among mixed gambles that are affected by loss aversion, even when all consequences are strictly positive.

These main ideas of TGPT had already been proposed and evaluated by Birnbaum and Zimmermann (1998, Appendix B, Model 2), along with certain other "loss aversion" theories of the so-called endowment effect. They rejected TGPT model as a descriptive account of judgments of highest buying and lowest selling prices, as well as the theory of Tversky and Kahneman (1991) and an anchoring and adjustment model.

In response to unpublished findings by Birnbaum and Yeary (1998) testing implications of configural weighting models, Luce (2000) developed a more elaborate theory in which prices and consequences are integrated via a joint receipt operation; this theory was further developed and evaluated by Birnbaum et al. (2016). The main difference between TGPT and Luce's (2000) approach is that in Luce's approach prices are integrated with consequences via a joint receipt operation rather than by simple addition or subtraction as in Birnbaum and Zimmermann (1998) and in Schmidt et al. (2008). However, in Luce's (2000) approach, like that of Birnbaum and Stegner (1979), the utility of negative consequences ("loss aversion") plays no role in the theory.

The present article presents new analyses of previously published data to evaluate the empirical status of TGPT. These results show that TGPT is not an accurate empirical description of either judgments of value (WTP and WTA) or of choices between prospects. The rest of this paper is organized as follows: Sect. 2 presents the key ideas of TGPT and presents theorems of TGPT (testable properties) that can be evaluated empirically; Sect. 3 presents evidence that these implications of TGPT are violated systematically by empirical findings. Section 4 presents a configural weight model (Birnbaum and Stegner 1979) and shows that it explains these phenomena that refute TGPT and provides a better fit to data. Section 5 discusses the implications of those empirical results and related findings for theories of choice and judgment.

2 Third-generation prospect theory: testable implications

Schmidt et al. (2008) based TGPT on CPT, which they called "second-generation" prospect theory. Let $G = (y_1, p_1; y_2, p_2; \ldots; y_m, p_m; x_n, q_n; \ldots; x_2, q_2; x_1, q_1)$ represent a gamble with outcomes ranked such that $x_1 < x_2 < \cdots < x_n < z \leq y_m < \cdots < y_2 < y_1$, where z is the status quo. Define cumulative probabilities of losses (worse than status quo) as follows: $Q_i = \sum_{k=1}^{i} q_k$; define decumulative probabilities of gains (better than or equal to status quo) as follows: $P_j = \sum_{k=1}^{j} p_k$. CPT (Tversky and Kahneman 1992) can be written as follows:

$$\mathrm{CPU}(G) = \sum_{i=1}^{n} \left[W^{-}(Q_i) - W^{-}(Q_{i-1})\right] u(x_i) + \sum_{j=1}^{m} \left[W(P_j) - W(P_{j-1})\right] u(y_j), \quad (1)$$

where Q_j and Q_{i-1} are the probabilities of a loss being equal to or worse (lower) than and strictly lower than x_i, respectively; P_j and P_{j-1} are the probabilities of winning a positive prize of x_j or better, and of strictly better than x_j, respectively ($P_0 = Q_0 = 0$). CPU(G) is the utility ("subjective value") of the gamble; the model assumes that $G \succ F \Leftrightarrow \text{CPU}(G) \succ \text{CPU}(F)$, where $\succ$ denotes systematic preference (apart from error). The functions $W(P)$ and $W^{-}(Q)$ are strictly increasing probability weighting functions, $W(0) = W^{-}(0) = 0$, and $W(1) = W^{-}(1) = 1$.

For binary mixed gambles of the form, $G = (y, p; x)$, where $y > z > x$, it is useful to note that

$$\text{CPU}(G) = \text{W}(p)u(y) + W^{-}(1 - p)u(x).$$

Utility (sometimes called "value") is defined with respect to changes from the status quo (z); it is often assumed that $z = 0$ and $u(0) = 0$.

In third-generation prospect theory (TGPT), CPT is retained for choices between prospects. However, for decisions to buy or sell, it is further assumed that the decision maker integrates the price of a prospect with the prizes. Thus, the price paid or received plays the role of the status quo, z, and the consequences of the gamble are compared to that value, which differs from prospect to prospect. Because the price paid (or demanded) depends on the gamble, it means that a positive cash prize might be either a gain or a loss, depending on the other features of each gamble.

For example, consider a binary gamble with nonnegative consequences, $G = (y, p; x)$, to win y with probability p and otherwise win x, where $y > x \geq 0$. In willingness to pay, it is assumed that the subject considers that if he/she pays buying price, B, and wins y, then the gain will be $y - B$, but if the gamble yields only x, then the loss will be $x - B$. Similarly, in willingness to accept, it is assumed that the subject considers a sale for S to be a gain when x occurs, since the profit is $S - x$; but the seller considers it a loss if the higher outcome y occurs, because the seller would have been better off to have kept the gamble; in this case, seller experiences a loss of $S - y$. Thus, buying and selling of prospects that yield strictly positive consequences involve the evaluation of mixed gambles, and psychological losses (e.g., for having sold a gamble that would have won) invoke utility and weighting of a loss.

These assumptions of TGPT (Schmidt et al. 2008, p. 209) lead to the following expressions for B (WTP) and S (WTA):

$$\text{CPU}(y - B, p; x - B) = 0 \quad (2)$$

$$\text{CPU}(S - y, p; S - x) = 0 \quad (3)$$

Thus, the price paid or received is integrated into the consequences of the gamble, producing a mixed gamble. Equations 1, 2, and 3 are termed TGPT. Equations 2 and 3 also appeared for maximal buying and minimal selling prices in Birnbaum and Zimmermann (1998, p. 178). [More general expressions were postulated by Luce (2000), in which subtraction is replaced by an (inverse) joint receipt operation. Luce's (2000) model is evaluated in Birnbaum et al. (2016).]

In the parameterized version of their model, Schmidt et al. (2008) also assume that utility can be approximated as follows:

$$u(x) = x^{\beta}, \qquad \text{for} \quad x \geq 0 \tag{4}$$

$$u(x) = -\lambda(-x)^{\beta}, \qquad \text{for} \quad x < 0, \tag{5}$$

where λ is a constant, sometimes called the “loss aversion” parameter. It is usually found that $\lambda > 1$ and that $0 < \beta < 1$, but such restrictions on these parameters are not necessary to what follows.

2.1 Complementary symmetry

Birnbaum and Zimmermann (1998) showed that Eqs. 1–5 imply a property they called *complementary symmetry* for binary gambles. The maximal buying price (WTP), B for gamble $(y, p; x)$ plus the minimal selling price, S(WTA) for the complementary gamble $(y, 1 - p; x)$ should be $x + y$.

Assuming Eqs. 1–5, with the definition, $T(p) = [\lambda W^{-}(1 - p)]/W(p)]^{(1/\beta)}$, it follows that

$$B = [y + T(p)x]/[1 + T(p)] \tag{6}$$

$$S = [x + T(p)y]/[1 + T(p)] \tag{7}$$

Adding Expressions 6 and 7, we have

$$S + B = x + y. \tag{8}$$

Complementary symmetry (Eq. 8) thus follows from Eqs. 1–5 with any weighting functions. Birnbaum and Zimmermann (1998) proved Eqs. 1–5 imply complementary symmetry for all gambles of the form $(y, 1/2, x)$; and Birnbaum et al. (2016) deduced Eqs. (6), (7), and (8) for all $(y, p; x)$. Michal Lewandowski (personal communication, April 23, 2016) reports that he has proved that TGPT (Eqs. 1, 2, 3) implies complementary symmetry for any $u(x)$ function on gains and losses.

2.2 First order stochastic dominance in judgments

Birnbaum (1997) devised a recipe that was tested in choice by Birnbaum and Navarrete (1998) who found that about 70% of undergraduates chose G = (\$96, 0.85;\$90, 0.05;\$12, 0.10) over F = (\$96, 0.90;\$14, 0.05;\$12, 0.05), even though F dominates G by first-order stochastic dominance. Birnbaum (1997, 2005) constructed this recipe to compare descriptive models that satisfy dominance against certain rival descriptive models that violate dominance. Because CPT must satisfy stochastic dominance, evidence that people systematically violate dominance shows that CPT is not an accurate descriptive model of risky decision making.

Stochastic dominance also follows in TGPT: judgments of buying or selling prices should also satisfy first-order stochastic dominance in this recipe; that is, the WTP (WTA) of F should exceed WTP (WTA) of G (Birnbaum et al. 2016).

2.3 Violations of restricted branch independence

Consider three-branch gambles with a fixed probability distribution. Let (x, y, z) represent a prospect to win x with probability p; y with probability q, and otherwise win z. Let $B(x, y, z)$ and $S(x, y, z)$ represent the judged value of WTP and WTA (highest buying and lowest selling prices of this prospect). Restricted branch independence can be expressed for such three-branch gambles as follows:

$$S(x, y, z) > S(x', y', z) \quad \text{if and only if} \quad S(x, y, z') > S(x', y', z') \tag{9a}$$

$$B(x, y, z) > B(x', y', z) \quad \text{if and only if} \quad B(x, y, z') > B(x', y', z') \tag{9b}$$

This property can be violated by TGPT when the weighting functions are not linear. The manner of violation, however, depends on the shape of the weighting functions. According to TGPT, it should be possible to predict the types of violations of this property in judgments from the shape of the weighting functions estimated from choices (Birnbaum 2008; Birnbaum and Zimmermann 1998).

For example, suppose $z' > y' > y > x > x' > z \geq 0$. There are two types of violations, Type 1: $B(x, y, z) > B(x', y', z)$ and $B(x, y, z') < B(x', y', z')$, or Type 2: $B(x, y, z) < B(x', y', z)$ and $B(x, y, z') > B(x', y', z')$. The same types stated here for buying prices (WTP) can be stated for selling prices (WTA). It can be shown that if the weighting functions are inverse-S in shape (as assumed by Tversky and Kahneman 1992; Schmidt et al. 2008, and others), then systematic violations should only be of Type 2 for either buying or selling prices. For example, the parameterized model of Schmidt et al. (2008) implies that B(\$39, \$45, \$2)< B(\$12, \$96, \$2) and B(\$39, \$45, \$148)> B(\$12, \$96, \$148). An analysis of restricted branch independence and its connection to the shape of the weighting function is presented in Birnbaum (2008, pp. 484–487). Intuitively, when the common branch (z) is the lowest, y' and y are highest-ranking values and have relatively greater weight than x' and x, which are middle values; but when the common branch is highest (z'), y' and y are intermediate-valued and have less weight than x' and x, which are the lowest valued consequences.

The weighting function of CPT required to reproduce standard findings in the literature must have this inverse-S form (Tversky and Kahneman 1992; Birnbaum 2008; Schmidt et al. 2008; Wakker 2011). Because TGPT requires the same weighting functions for both choice and for judgments of value, it predicts the Type 2 pattern of violation of restricted branch independence in choice tasks and in WTP and WTA judgments.

3 Empirical violations of the implications of TGPT

3.1 Violations of complementary symmetry

In Sect. 2, it was shown that in TGPT, the sum of buying and selling price of complementary binary gambles of the form $(y, p; x)$ should equal the sum of the outcomes. This sum should, therefore, be independent of other factors, such as range, $|x - y|$,

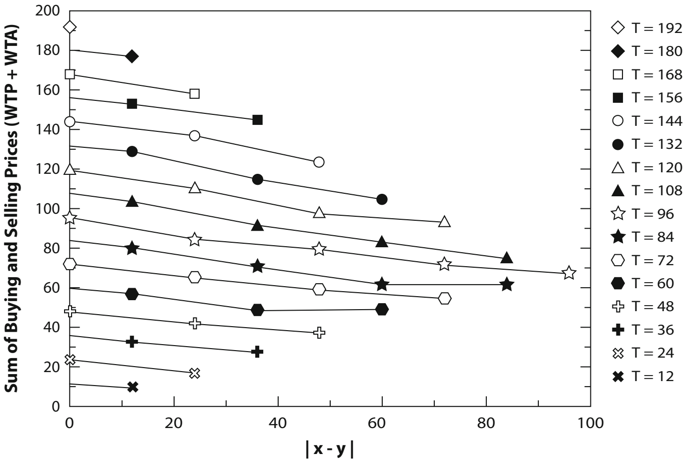

Fig. 1 Sum of median WTP and WTA (buying plus selling prices) for gambles of the form $(y, 0.5; x)$ as a function of $|x - y|$ with a separate curve for each level of $T = x + y$. According to complementary symmetry, *all curves* should be *horizontal*. Instead, all curves decrease as a function of range

holding p and $x + y$ constant. However, empirical evidence contradicts this implication, as shown next.

At the time of Birnbaum and Sutton (1992), neither TGPT nor the property of complementary symmetry had yet been developed; nevertheless, that experiment provides tests of that property. Figure 1 shows a reanalysis of data from Birnbaum and Sutton (1992) to test complementary symmetry. The figure shows the sum of median judgments of buying and selling prices (WTP + WTA) of gambles of the form $(y, 0.5; x)$; that is $S + B$ of Eq. 8. These are plotted as a function of $|x - y|$ with a separate curve for each value of $T = x + y$. According to TGPT, each curve should be horizontal with a constant value of $x + y$.

In contradiction to TGPT, Fig. 1 shows that $S + B$ decreases systematically with range for every value of T. For example, the median WTP and WTA of ($60, 0.5; $48) are $50 and $54, respectively, for a total of $104. However, the median WTP and WTA of ($96, 0.5; $12) are $25 and $50, respectively, for a total of only $75. TGPT implies that both totals should have been $108. For all 36 points in Fig. 1 where the range is not zero, $S+B < x+y$. Furthermore, every curve decreases as a function of range ($|x-y|$). Because there are 15 curves in Fig. 1 connecting at least two empirical points, the probability that the right most point in each curve would fall below the left most point is one-half to the 15th power, assuming complementary symmetry. Therefore, these data systematically violate complementary symmetry, contrary to TGPT. Other violations of complementary symmetry are reported in Birnbaum et al. (2016), who showed that holding $x + y$ and $|x - y|$ constant, $S + B$ varies systematically as a function of p.

3.2 Violations of first-order stochastic dominance

Birnbaum and Yeary (1998) asked 66 undergraduates to evaluate 166 risky gambles from the viewpoints of both highest buying price and lowest selling price. Interspersed

Table 1 Median Judgments in the tests of stochastic dominance

Test	Buying prices (WTP)			
	$G+$	Median WTP	$G-$	Median WTP
1	0.05 0.05 0.90 / \$12 \$14 \$96	30.0	0.10 0.05 0.85 / \$12 \$90 \$96	60.0
2	0.06 0.06 0.88 / \$3 \$5 \$97	22.5	0.12 0.04 0.84 / \$3 \$92 \$97	50.0
3	0.02 0.03 0.95 / \$6 \$8 \$99	40.0	0.05 0.03 0.92 / \$6 \$91 \$99	54.0
4	0.01 0.01 0.98 / \$4 \$7 \$97	50.0	0.02 0.02 0.96 / \$4 \$89 \$97	62.5
Test	Selling prices (WTA)			
	$G+$	Median WTA	$G-$	Median WTA
1	0.05 0.05 0.90 / \$12 \$14 \$96	73.5	0.10 0.05 0.85 / \$12 \$90 \$96	81.5
2	0.06 0.06 0.88 / \$3 \$5 \$97	68.0	0.12 0.04 0.84 / \$3 \$92 \$97	80.0
3	0.02 0.03 0.95 / \$6 \$8 \$99	82.5	0.05 0.03 0.92 / \$6 \$91 \$99	83.5
4	0.01 0.01 0.98 / \$4 \$7 \$97	81.0	0.02 0.02 0.96 / \$4 \$89 \$97	87.5

In each test, $G+$ dominates $G-$ and yet receives a lower median judgment

among these trials were eight trials that provided four tests of first-order stochastic dominance (FOSD) in each point of view. These four tests of FOSD were constructed to show violations, according to configural weight models. Table 1 shows the median judgments of WTP and WTA for these 8 gambles.

In all eight comparisons in Table 1 (four tests by two viewpoints), the dominated gamble (denoted $G-$ in Table 1, to the right) received higher median judgments than the dominant gamble (G+, to the left), violating FOSD. Means are similar and show the same violations (Birnbaum et al. 2016). The overall mean judgment was \$63.31 for dominated gambles (averaged over the four tests and two viewpoints), compared to a mean of \$55.11 for dominant gambles. This difference between dominant and dominated gambles is significant; therefore, the null hypothesis that the dominance relation has no effect on judgments of value was rejected in favor of the hypothesis that people assign higher judgments to dominated gambles in these specially constructed pairs of gambles.

Analysis of individuals' data showed that 51 of 66 participants (77%) assigned higher mean judgments to dominated gambles than to the dominant gambles, averaged over the four tests and two viewpoints.

These results with judgment agree with results from direct choices between the same pairs of gambles; Birnbaum and Navarrete (1998) found that 73, 61, 73, and 73 judges (out of 100) chose the dominated gamble $G-$ over the dominant gamble, $G+$, in direct choices of Tests 1 though 4 (of Table 1), respectively. Many subsequent studies have confirmed high rates of violation of FOSD in direct choices constructed from this recipe and shown that they are not due to random error (Birnbaum and Bahra 2012). In sum, violations of FOSD are observed in buying prices, selling prices, and in direct choices. All three results are systematic violations of TGPT.

Because FOSD follows in CPT for choice and TGPT for both WTP and WTA for any utility and any decumulative weighting functions, it would not be possible to salvage these models as descriptive of such violations by choosing other parameters or other functions for u, W, and W^-.

3.3 Violations of restricted branch independence

Because TGPT assumes that the same $W^-(Q)$ and $W(P)$ functions apply to judgments as to choices, this theory implies that we should be able to predict violations of restricted branch independence from the shape of the probability weighting functions estimated from choice experiments. It is well-known that to describe standard results of empirical choice studies, including the Allais paradoxes, CPT requires inverse-S probability weighting functions in which intermediate branches receive lower weight than lowest or highest valued branches (Tversky and Kahneman 1992; Birnbaum 2008; Wakker 2011).

Empirically, the observed type of violation of restricted branch independence in both WTP and WTA judgments is not in agreement with the inverse-S weighting function postulated in TGPT (Birnbaum and Beeghley 1997; Birnbaum and Zimmermann 1998; Birnbaum and Veira 1998). In these studies, it is typically found that the violations are of Type 1 rather than Type 2 (Sect. 2.3).

For example, Table 2 shows mean judgments of 12 of the prospects studied by Birnbaum and Beeghley (1997), who asked 46 participants to judge both WTP and WTA for 166 gambles, each of which had three, equally likely outcomes: (y, x, z). The mean (and median) judgments violate restricted branch independence in both viewpoints in the opposite way from that predicted by TGPT.

The predicted judgments in Table 2 are calculated from TGPT using the parameters estimated by Tversky and Kahneman (1992). The observed type of violations of restricted branch independence are opposite of the type predicted. Note that in both WTA and WTP viewpoints, the empirical violations are of Type 1 rather than Type 2; for example, $B(x, y, z) > B(x', y', z)$ and $B(x, y, z') < B(x', y', z')$, as well as $S(x, y, z) > S(x', y', z)$ and $S(x, y, z') < S(x', y', z')$, when $z' = \$148$, $y' = \$96$, $y = \$45$, $x = \$39$, $x' = \$12$, and $z = \$2$. In contrast, TGPT predicts the opposite orderings from these empirical results.

When the common consequence is $z = \$2$ (upper half of the table), TGPT predicts that $(x', y', z) = (\$2, \$12, \$96)$ should receive the highest judgment in both WTP and WTA compared to the other values of $(\$2, x, y)$ in the upper half of Table 2. Instead, (\$2, \$12, \$96) is judged lowest in WTP and it falls third from the bottom in WTA. Similarly, TGPT predicts that $(x', y', z') = (\$12, \$96, \$148)$ should receive the lowest judgment for WTP and fourth from the bottom for WTA compared to the other values of $(x, y, \$148)$ in the lower half of the table, but empirically it falls third highest for WTP and highest for WTA.

In sum, the data show systematic Type 1 violations, whereas TGPT with an inverse-S weighting function implies that the violations should be of Type 2.

Violations of restricted branch independence in direct choice also show the opposite pattern from that predicted by the inverse-S weighting function needed by CPT and

Table 2 Reanalysis of data from Birnbaum and Beeghley (1997)

Lottery	WTP	WTA	Pred WTP	Pred WTA
($2, $27, $33)	15.4	23.0	6.7	27.0
($2, $33 $39)	19.1	26.6	7.7	33.0
($2, $39, $45)	19.6	30.0	8.6	38.9
($2, $45, $51)	21.9	34.2	9.6	44.9
($2, $51, $57)	27.7	37.1	10.5	50.9
($2, $12, $96)	14.4	28.5	11.9	61.5
($27, $33, $148)	35.5	51.9	38.6	102.7
($33, $39, $148)	39.8	50.2	44.1	105.0
($39, $45, $148)	45.2	58.5	49.6	107.3
($45, $51, $148)	49.9	62.0	55.0	109.6
($51, $57, $148)	56.5	68.5	60.5	111.8
($12, $96, $148)	47.8	75.2	30.7	108.0

Predicted WTP and WTA are based on third-generation prospect theory, using parameters of Tversky and Kahneman (1992)

TGPT. This pattern of violations (Type 1) has been replicated in dozens of empirical studies using different formats for presentation of choices (Birnbaum 2004, 2008; Birnbaum and Bahra 2012; Birnbaum and Navarrete 1998). Therefore, one cannot retain both TGPT and the inverse-S decumulative weighting function, if one wants to describe empirical data for WTP, WTA, or choice.

4 Configural weighting models

In social and evaluative judgment tasks, it has been found that unfavorable or negative information seems to override the effects of positive or favorable information. Birnbaum (1974) considered two theories to explain this effect: either the lower-valued information has more extreme value (analogous to "loss aversion") or it has greater configural weight. In a series of experiments, Birnbaum (1974, 1982) concluded that the lower-valued or negative information carries greater configural weight. Birnbaum represented the combination of information by an averaging model with weights that depend on ranks.

Although Birnbaum (1974) wrote of weights that "depend on the ranks" of the components, such configural weighting differs from what later came to be called "rank-dependent weighting", as described in Quiggin (1982, 1993). The term "rank-dependent" weighting now refers to weighting in which cumulative weight is a monotonic function of cumulative probability. In Birnbaum's configural weighting, however, the weight of each branch (e.g. discrete probability-consequence pair presented to the decider) is affected by the rank of its discrete consequence, but not necessarily a function of cumulative probability.

Birnbaum and Stegner (1979, Experiment 5) theorized that buying and selling prices would induce different patterns of configural weighting. An extension of Birnbaum and Stegner's model is now known as the transfer of attention exchange (TAX) model. It has proven useful for describing choices between risky prospects and it correctly predicted results with a dozen "new paradoxes" that cannot be described by CPT (Birnbaum 2008).

For gambles of the form, $G = (y_1, p_1; y_2, p_2; \ldots; y_k, p_k; \ldots; y_i, p_i; \ldots; y_n, p_n)$ where $0 \leq y_n < \cdots < y_i < \cdots < y_k < \cdots < y_2 < y_1$, the TAX model can be written:

$$\text{TAX}U(G) = \frac{\sum_{i=1}^{n} t(p_i)u(y_i) + \sum_{i=1}^{n}\sum_{k=1}^{i}[u(y_i) - u(y_k)]\omega(p_i, p_k, n)}{\sum_{i=1}^{n} t(p_i)}, \quad (10)$$

where $\omega(p_i, p_k, n)$ represents the weight transferred from branch k to branch $i (k \leq i$; hence weight is transferred from branches with higher-valued consequences to branches with lower consequences when $\omega(p_i, p_k, n) > 0)$.

The "special" TAX model is a special case in which all weight transfers are the same fixed proportion of the weight of the branch giving up weight, as follows:

$$\omega(p_i, p_k, n) = \begin{bmatrix} \frac{\delta \cdot t(p_k)}{n+1}, & \delta > 0 \\ \frac{\delta \cdot t(p_i)}{n+1}, & \delta \leq 0 \end{bmatrix} \quad (11)$$

In this special case of the TAX model, the amount of weight transferred between any two branches is a fixed proportion of the (transformed) probability of the branch losing weight. If lower-ranked branches have more importance (as they would for a "risk-averse" person), it is theorized that weight is transferred from branches with higher consequences to those with lower-valued consequences, i.e., $\delta > 0$. The term "prior TAX" is used to refer to specific parameters in special TAX, where $u(x) = x$, $t(p) = p^{0.7}$, and $\delta = 1$. These parameters are not optimal (selected in 1996) but were fairly successful in predicting modal choices by undergraduates in new studies done over the next two decades (Birnbaum 2008; Birnbaum and Bahra 2012) , who made choices between risky prospects involving small positive consequences (less than \$150).

For the case of binary, 50–50 gambles, $G = (y, 0.5; x)$, special TAX (Eqs. 10, 11) further simplifies to

$$\text{TAX}(y, 0.5; x) = 0.5[u(y) + u(x)] + \omega|u(x) - u(y)|, \quad (12)$$

where $\omega = -\delta/6$. Note that this holds for any $t(p)$ function.

Note that the weight of the higher ranked utility is $0.5 + \omega$, and the weight of the lower ranked value is $0.5 - \omega$. If $\omega = 0$, the model is a simple average (EU for 50–50 gambles), but when ω is positive, the higher valued stimulus gets greater weight, and when ω is negative, the lower valued stimulus gets greater weight. At the extremes, when $\omega = 0.5$, or $\omega = -0.5$,the model becomes a maximum or minimum model.

Birnbaum and Stegner (1979, Experiment 5) theorized that configural transfers of weight (values of ω) would be affected by the judge's point of view, which they theorized were due to psychological incentives to a judge for over- versus under-estimating value (deduced in Birnbaum et al. 1992). To test their theory, they asked people to judge the most a buyer should pay, the least a seller should accept, or the "fair" price of used cars, based on evaluations provided by people of varied bias and expertise who had examined the cars. Their theory (Birnbaum and Stegner 1979, p. 60–61) was that buyers, sellers, and independents (judging "fair" value) would have different values of ω, but $u(x)$ is independent of point of view. This model fit the data well, and correctly predicted the finding that WTP and WTA are not monotonically related to each other (Birnbaum and Stegner 1979; Birnbaum 1982; Birnbaum and Zimmermann 1998).

Birnbaum and Stegner (1979) estimated the configural weights for buyers, sellers, and "fair price" judgments as follows: $\omega_B = -0.19$ (buyers put greater weight on lower values), $\omega_S = 0.06$ (sellers put greater weight on higher values), and $\omega_F = -0.07$ ("fair" prices fall intermediate between buyer and seller).

4.1 Configural weighting account of violations of complementary symmetry

To predict violations of complementary symmetry for gambles of the form $(y, 0.5; x)$ one can assume Eq. 12 holds for both S and B, with the assumption that $u(x) = x$ and that only ω differs for S and B:

$$S(x, 0.5; y) + B(y, 0.5; x) = x + y + (\omega_B + \omega_S)|x - y|,$$

where ω_B and ω_S are the configural weights for buying and selling prices, WTP and WTA, respectively. If $\omega_B = -0.19$ and $\omega_S = 0.06$, as found by Birnbaum and Stegner (1979), then $\omega_B + \omega_S = -0.13$; therefore, one would predict a decline in Fig. 1 as a function of $|x - y|$ for each $T = x + y$.

Using parameter estimates from Birnbaum and Stegner (1979) for used cars is obviously not optimal for fitting data of Birnbaum and Sutton (1992) for risky prospects, but it does show that the general trend is consistent with the theory and that a decline is anticipated, given parameters fit to previous data. Thus, the configural weight model can violate complementary symmetry and can correctly predict the general trends in Fig. 1.

4.2 Violations of stochastic dominance in configural weight models

Although Birnbaum (1974) had used the terminology that "weights depend on ranks" to describe configural weight models such as Eq. 10, the configural weight models do not in general imply stochastic dominance, as implied by "rank dependent utility" models as in Quiggin (1993), Luce and Fishburn (1991) or Tversky and Kahneman (1992). To compare configural weight models against these rank-dependent models, Birnbaum (1997) derived critical tests to distinguish them. Among these tests was a recipe in which the configural weight models were expected to violate first-order stochastic dominance.

To understand how and when the configural weight models (as in Eqs. 10, 11) imply violations of FOSD, realize that when $t(p)$ is a negatively accelerated function of p, splitting a branch can produce splinters having greater total weight than the unsplit branch.

Birnbaum's (1997) recipe is illustrated as follows: starting with $G0 = (\$96, 0.9; \$12, 0.1)$, split the upper branch and reduce the value on the splinter to create a strictly worse gamble $G- = (\$96, 0.85; \$90, 0.05; \$12, 0.10)$; now split the lower branch of $G0$ and increase the value of the splinter to create a strictly better gamble, $G+ = (\$96; 0.90; \$14, 0.05; \$12, 0.05)$. By splitting the upper branch to create $G-$, the sum of the weights of the two upper branches has increased (because $t(p)$ is negatively accelerated), thus improving the value of the gamble (even though it has been made objectively worse); similarly, splitting the lower branch of a gamble makes it seem worse, so $G+$ has a lower value despite being objectively better than $G0$.

For example, with $t(p) = p^{0.7}$, $u(x) = x$, and $\delta = 1$, the prior TAX model values of $G+$ and $G-$ are \$45.77 and \$63.10, respectively; therefore, TAX predicts a violation of first order stochastic dominance in this case. Birnbaum et al. (2016) showed that with plausible parameters for buying and selling prices, the TAX model can accommodate violations of FOSD like those in Table 1.

4.3 Violations of restricted branch independence

Birnbaum (2008) presented an analysis of violations of restricted branch independence in the special TAX model (Eqs. 10, 11) and showed that if violations occur, they must be of Type 1. If $\delta = 0$, then there will be no violations of restricted branch independence. However, whether weight is transferred from highest valued branches to lower valued branches, or vice versa, TAX implies any systematic violations must be of Type 1 (see Birnbaum 2008, Fig. 11).

Birnbaum and Beeghley (1997) fit Eq. 10 to their judgments, including those in Table 2. The weights of lowest, middle, and highest of three equally likely consequences were estimated to be 0.56, 0.36, and 0.08 in the buyer's viewpoint (WTP), respectively, and they were 0.27, 0.52, and 0.21 in the seller's viewpoint (WTA), respectively. In both cases, the middle valued branch does not have the least weight, contrary to the inverse-S weighting function of CPT. As shown in Birnbaum and Beeghley, these weights do an excellent job of fitting the violations of restricted branch independence and the non-monotonic relationship between WTP and WTA (including Table 2), even with the assumption that $u(x) = x$.

4.4 Model fitting: comparison of fit

The TGPT models of WTP and WTA (Eqs. 6, 7) were fit to judgments of 63 binary gambles of the form, $(y, p; x)$ by Birnbaum et al. (2016). There were 9 levels of probability: $p = 0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95$, or 0.99; there were 7 levels of (y, x): $(y, x) = (\$100, \$0), (\$72, \$0), (\$48, \$0), (\$24, \$0), (\$100, \$6), (\$100, \$24)$ or $(\$100, \$48)$. In this case, 9 parameters were estimated for $T(p)$ in Eqs. 6 and 7, because there were 9 levels of p. Estimating $T(p)$ for each p allows complete flexibility to the weighting functions, W and W^- so they need not follow the inverse-S shape or any particular form in this analysis; λ and β were also free in Eqs. 4 and 5, so there were 11 free parameters.

Despite the flexibility allowed by so many free parameters, TGPT does not fit the data as well as configural weight models that used fewer parameters. The sum of squared deviations between predicted and obtained judgments (126 predicted values for 63 gambles in WTP and WTA) was 20,242 for TGPT (11 parameters) compared to 1051 for the TAX model with 6 parameters and 1097 for a TAX model with 5 free parameters (where $u(x) = x$). This comparison shows how much better the older, configural weight models performed in fitting judgments.

An anomaly not featured in this paper also violates TGPT: violations of consequence monotonicity. Increasing the lowest consequence in a binary gamble, holding everything else fixed, should strictly improve judgments, but it has been found when a low probability consequence is reduced from a small positive value to zero, it can actually increase WTP and WTA (Birnbaum and Sutton 1992; Birnbaum et al. 1992) This anomaly has been described by the assumption that the value zero (status quo) receives lower weight than nonzero consequences (Birnbaum 1997). These violations of monotonicity as well as violations of complementary symmetry contributed to the difference in fit between TAX and TGPT in Birnbaum et al. (2016). Additional details of the experiment and its analysis, as well as evaluations of other models including that of Luce (2000), are presented in Birnbaum et al. (2016).

5 Discussion and conclusions

This paper analyzed three properties that are implied by TGPT to show that they are violated by empirical data: complementary symmetry, first-order stochastic dominance, and the Type 2 violation of restricted branch independence. It also noted that when the models are fit to data from a factorial experiment of binary gambles, the judgments of WTP and WTA are better fit by configural weight models than by TGPT.

Some articles on the endowment effect report the ratio of WTA/WTP, as if this is a sensible or stable index. According to both TGPT and configural weight models, this ratio is not an invariant and can be manipulated at will. Empirically, the ratio of WTA/WTP = 54/50 = 1.08 for the gamble Q = ($60, 0.5; $48) and WTA/WTP = 50/25 = 2 for the gamble R = ($96, 0.5; $12). Note that the ranges differ ($12 versus $84), so TAX predicts this effect. Not only is this ratio not a constant, but WTA is not even a monotonic function of WTP (Birnbaum 1982; Birnbaum and Stegner 1979; Birnbaum and Sutton 1992). For example, for gamble U = ($36, 0.5; $48), WTP = $39.5 and WTA = $41. Thus, for WTP, U is rated higher than R ($39 > $25) but for WTA, U is rated lower than R ($41 < $50). As noted by Birnbaum et al. (2016), such preference reversals between buying and selling prices refute theories in which WTA is a monotonic function of WTP or proportional to it, as in Tversky and Kahneman (1991).

Because TGPT relies on CPT, it is also refuted by studies of choice testing Birnbaum's (2008) "new paradoxes". These new paradoxes represent findings that refute prospect theories in the same way that the Allais paradoxes refuted EU: modal data patterns lead to self-contradiction within CPT. Such paradoxes refute CPT as a descriptive theory of choices between gambles. Several of these new paradoxes that refute CPT have been replicated in dozens of experiments with a variety of experimental procedures.

For example, the violations of first-order stochastic dominance reported here in Table 1 have also been found in more than 40 studies of choice with probability represented by numbers of balls of different colors in urns, by frequencies of tickets with different prize values printed on them, with pie charts representing spinners, with bar charts showing probabilities, with lists of equally likely outcomes, with independent and dependent gambles, with decumulative probabilities, and with different arrangements of juxtaposing branches in the gambles compared. They have been observed in lab studies of hypothetical choice, in Internet studies with chances of real prizes, and in public settings where real cash prizes are awarded in the presence of an excited group via drawings conducted immediately after the choices (reviewed in Birnbaum 2008; Birnbaum and Bahra 2012).

The fact that violations of stochastic dominance and restricted branch independence yield such similar results in both judgments and in direct choices suggests that these violations arise from a common evaluation mechanism for the evaluation of gambles, rather than from some mechanism specific to comparison of gambles.

In some historical instances, anomalies have been found that violate older models and newer models were created to accommodate those new findings. It is worth emphasizing, however, that in this case, older models were used to devise new tests of prospect theories. The configural weight models were used to design new tests of critical properties that would distinguish between them and the newer models, leading to the “new paradoxes” and “anomalies” that refute both CPT and TGPT. In the case of complementary symmetry, a new test was deduced that could be evaluated using older data as well as the older theory.

Cognitive mechanisms that might underlie configural weighing have been explored by Johnson and Busemeyer (2005) and by Ashby et al. (2012).

The configural weight models can explain the finding that the ratio of WTA to WTP is not a constant and that these two judgments are not even monotonically related to each other (Birnbaum and Stegner 1979). It correctly predicted the trend of violations of complementary symmetry in Fig. 1 (decreasing as a function of increasing range), it correctly describes the type of violations of restricted branch independence (Type 1), it was used to devise the recipe for violations of first-order stochastic dominance that have been observed, and it fits data better than the parameterized model of Schmidt et al. (2008).

It seems reasonable that investigators who pursue the concept of loss aversion as a theory of the endowment effect should be asked to show that their loss aversion models provide better fits to the data than provided by the earlier models. Perhaps those working with the concept of “loss aversion” can find a new theory to account for empirical phenomena such as these violations of complementary symmetry, first-order stochastic dominance, and restricted branch independence.

References

Ashby, N. J. S., Dickert, S., & Glöckner, A. (2012). Focusing on what you own: Biased information uptake due to ownership. *Judgment and Decision Making*, *7*, 254–267.

Birnbaum, M. H. (1974). The nonadditivity of personality impressions. *Journal of Experimental Psychology Monograph*, *102*, 543–561.

Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 401–485). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Birnbaum, M. H. (1997). Violations of monotonicity in judgment and decision making. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 73–100). Mahwah, NJ: Erlbaum.

Birnbaum, M. H. (2004). Tests of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effects of format, event framing, and branch splitting. *Organizational Behavior and Human Decision Processes*, *95*, 40–65.

Birnbaum, M. H. (2005). A comparison of five models that predict violations of first-order stochastic dominance in risky decision making. *Journal of Risk and Uncertainty*, *31*, 263–287.

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review*, *115*, 463–501.

Birnbaum, M. H., & Bahra, J. P. (2012). Separating response variability from structural inconsistency to test models of risky decision making. *Judgment and Decision Making*, *7*, 402–426.

Birnbaum, M. H., & Beeghley, D. (1997). Violations of branch independence in judgments of the value of gambles. *Psychological Science*, *8*, 87–94.

Birnbaum, M. H., Coffey, G., Mellers, B. A., & Weiss, R. (1992). Utility measurement: Configural-weight theory and the judge's point of view. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 331–346.

Birnbaum, M. H., & Navarrete, J. B. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty*, *17*, 49–78.

Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, *37*, 48–74.

Birnbaum, M. H., & Sutton, S. E. (1992). Scale convergence and utility measurement. *Organizational Behavior and Human Decision Processes*, *52*, 183–215.

Birnbaum, M. H., & Veira, R. (1998). Configural weighting in judgments of two- and four-outcome gambles. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 216–226.

Birnbaum, M. H., Yeary, S. (1998). *Tests of Stochastic Dominance, Cumulative Independence, Branch Independence, Coalescing, Event-Splitting Independence, and Asymptotic Independence in Buying and Selling Prices.* Unpublished manuscript.

Birnbaum, M. H., Yeary, S., Luce, R. D., & Zhao, L. (2016). Empirical evaluation of four models for buying and selling prices of gambles. *Journal of Mathematical Psychology*, *75*, 183–193.

Birnbaum, M. H., & Zimmermann, J. M. (1998). Buying and selling prices of investments: Configural weight model of interactions predicts violations of joint independence. *Organizational Behavior and Human Decision Processes*, *74*(2), 145–187.

Edwards, W. (1954). The theory of decision making. Psychological Bulletin, 51, 380–417.

Johnson, J. B., & Busemeyer, J. R. (2005). A dynamic, stochastic, computational model of preference reversal phenomena. *Psychological Review*, *112*, 841–861.

Luce, R. D. (2000). *Utility of gains and losses: Measurement-theoretical and experimental approaches*. Mahwah, NJ: Erlbaum. Errata: see http://www.imbs.uci.edu/files/personnel/luce/luce.html.

Luce, R. D., & Fishburn, P. C. (1991). Rank- and sign-dependent linear utility models for finite first order gambles. *Journal of Risk and Uncertainty*, *4*, 29–59.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, *3*, 324–345.

Quiggin, J. (1993). *Generalized expected utility theory: The rank dependent model*. Boston: Kluwer Academic.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*, 7–59.

Schmidt, U., Starmer, C., & Sugden, R. (2008). Third-generation prospect theory. *Journal of Risk and Uncertainty*, *36*, 203–223.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, *1*, 39–60.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics*, *106*(4), 1039–1061.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.

Wakker, P. (2011). *Prospect theory: For risk and ambiguity*. Cambridge, UK: Cambridge University Press.