Outside Good Uncertainty

Ordinal Dual Response in Choice-Based Conjoint Analysis

Prachi Bhalerao Dan Yavorsky Geoffery Zheng

2026-07-10

We provide a behavioral model in which consumers can identify their most-preferred option among a set of alternatives but face uncertainty in their evaluation of the outside good. The model rationalizes the dual-response format used in choice-based conjoint analysis and extends the follow-up purchase question from binary to ordinal, accommodating scales with any number of points. The resulting likelihood couples a multinomial logit for the forced choice with a cumulative-link ordinal model driven by the task's inclusive value, and it nests the leading binary dual-response models as its two-point special case. We derive the information gains from finer response scales, provide estimation routines for aggregate and hierarchical Bayesian versions of the model — including an extension with respondent-specific cut points — and demonstrate the model's superior performance relative to the dichotomization and constant-probability heuristics currently used by practitioners. An empirical application uses a choice-based conjoint study conducted in partnership with a consumer insights consultancy.

1 Introduction

Consumers are good at telling us which product they like best. They are considerably less certain about whether they would actually buy it. A shopper can pick a preferred bluetooth speaker from a shelf of a dozen — or a preferred color from a dozen options of the same model — and still hesitate at the moment of purchase; industry trackers consistently estimate that roughly seven in ten online shopping carts are abandoned before checkout.[1] The comparison among displayed alternatives and the commitment to purchase are different cognitive acts: the former is a choice among fully described options in front of the consumer, while the latter requires her to weigh the winner against an outside option — keep the money, wait for a sale, buy elsewhere, do without — whose value she has not yet fully resolved. Survey respondents

[1] TODO: add citable source. The Baymard Institute maintains a running meta-analysis of cart-abandonment studies with an average of roughly 70%.

reflect this distinction naturally. Asked "which of these would you choose?", they answer readily; asked "would you buy it?", many prefer to answer "probably" than to commit to yes or no. Uncertainty about a purchase *feels natural*, and stated purchase-likelihood scales have been a fixture of marketing research for over half a century (Juster 1966; Morrison 1979).

Choice-based conjoint analysis, the workhorse of stated-preference measurement in marketing (Allenby et al. 2019), has developed its own device for separating these two acts: the *dual-response* design. In each task the respondent first makes a forced choice among the displayed product profiles, and then answers a follow-up question about whether she would actually purchase her selection. The design was introduced to solve a practical problem with the traditional no-choice option. When "none of these" appears alongside the profiles in a single-stage task, every task in which it is chosen contributes nothing to the estimation of attribute preferences, and the efficiency loss grows with no-choice incidence (Brazell et al. 2006). The dual response recovers that lost information — the forced choice always reveals preference among profiles — while still measuring demand for the category through the follow-up question, and it does so without systematically biasing preference estimates (Brazell et al. 2006; Dhar and Simonson 2003). Following its implementation in Sawtooth Software's widely used CBC system (Diener et al. 2006; Sawtooth Software 2017), the dual-response format became standard practice in commercial conjoint research, and it has continued to attract methodological attention, from willingness-to-pay measurement (Schlereth and Skiera 2017) to the recent behavioral and econometric re-examinations of Hung, Liu, et al. (2025) and Hung, Kurz, et al. (2025).

Practice, however, has quietly moved beyond the models available to analyze it. The canonical dual-response question is binary — "would you buy the option you selected above? (yes/no)" — and the canonical models are correspondingly binary: a multinomial logit for the forced choice paired with a binary logit for the purchase question (Diener et al. 2006; Hung, Liu, et al. 2025). But many applied researchers instead field an *ordinal* follow-up — "how likely are you to buy the option you selected above?", answered on a five- or seven- or eleven-point scale from "very unlikely" to "very likely." The ordinal format is attractive precisely because purchase uncertainty feels natural to respondents: it lets them express a degree of conviction rather than forcing a resolution they have not reached. It is also, as we show, more informative. Yet there is no statistical model for it. Analysts confronted with ordinal dual-response data today either *dichotomize* the scale at an arbitrary point (top box or top two boxes count as "buy") and estimate the binary model, or apply *constant-probability weights* that map each scale point to a fixed purchase probability regardless of what was on the screen. The first discards information the researcher paid to collect and hard-codes an arbitrary cut; the second, as our model makes precise, ignores the fact that the meaning of "somewhat likely" depends on how attractive the displayed alternatives were.

This paper provides the missing model. We start from a simple behavioral premise: the consumer can rank the goods in front of her because she observes everything relevant about them, but she faces genuine uncertainty about the outside good — the not-yet-arrived purchase occasion that the follow-up question asks her to imagine. This premise does three things at

once. It *rationalizes the dual-response design*: the forced choice is the question the consumer can answer with certainty, and the follow-up is the question about which she is uncertain, so the design cleanly partitions what she knows from what she does not. It *rationalizes the ordinal format*: a consumer who has not resolved her outside-good uncertainty holds a purchase *probability*, and an ordinal scale is exactly an instrument for reporting a probability in bins. And it *delivers a likelihood*: because the maximum of Gumbel utilities is itself Gumbel — with location equal to the log-sum inclusive value — and is independent of which alternative attains it, the joint probability of the pair (chosen profile, ordinal report) factors into a multinomial logit times a cumulative-link ordinal model driven by the inclusive value of the task.

Within this framework we develop two variants that differ only in how literally the respondent carries her uncertainty into the report. In Model A she reports the purchase probability itself, binned by cut points on the unit interval; the implied ordinal likelihood has a Gumbel (complementary log-log type) link. In Model B she behaves as the binary literature implicitly assumes — resolving the outside-good draw and grading the resulting utility difference — which yields a standard ordered logit link. The two variants have identical parameter counts and are numerically close, but they carry different behavioral interpretations, and their binary special cases differ in an instructive way. With a two-point scale, Model B collapses *exactly* to the "Unified Model" of Hung, Liu, et al. (2025) — equivalently the DR-AnyMax model of Diener et al. (2006), and, after telescoping, the standard single-stage multinomial logit with a no-choice constant (Haaijer et al. 2001). Our model is thus a strict generalization of the leading binary dual-response model: practitioners who adopt it lose nothing at $W = 2$ and gain the ability to use every scale point of a richer follow-up. Model A's binary case, by contrast, implies that even the familiar yes/no follow-up should be analyzed with a complementary log-log rather than logistic link when respondents remain genuinely uncertain — a one-line modification of existing practice that is testable on existing data.

We make four contributions. First, the behavioral model itself: a utility-theoretic account of dual-response conjoint in which outside-good uncertainty is the primitive, which rationalizes the design, selects the link function, and nests the existing binary models as the two-point special case. Second, an analysis of the information content of the ordinal follow-up: we derive the Fisher-information increment contributed by a W-point second response and show it is positive semi-definite and non-decreasing in scale refinement, extending the binary efficiency results of Brazell et al. (2006) and Hung, Liu, et al. (2025) and giving design guidance on the choice of W. Third, estimation machinery: maximum likelihood for the aggregate model and a custom Markov chain Monte Carlo sampler for the hierarchical Bayesian model with heterogeneous preferences — including an extension with respondent-specific cut points that accommodates individual differences in scale usage (Greene and Hensher 2010; Falco et al. 2015) — with complete replication code in R. Fourth, evidence: a simulation study demonstrating parameter recovery and quantifying the losses from dichotomization and constant-probability weighting, and an empirical application to a commercial choice-based conjoint study with an ordinal dual response conducted in partnership with a consumer insights consultancy.

Our work sits at the intersection of three literatures. The dual-response and no-choice literature

in marketing established the design's efficiency rationale and its estimation conventions (Brazell et al. 2006; Diener et al. 2006; Haaijer et al. 2001; Schlereth and Skiera 2017; Hung, Liu, et al. 2025), and recent work shows the outside good remains the format's active research frontier — including evidence that its utility can drift over the course of a survey (Hung, Kurz, et al. 2025). The opt-out literature in choice modeling catalogs how misspecifying the outside option distorts welfare and share predictions (Campbell and Erdem 2019). And a literature on ordered and mixed response formats in preference measurement — from unified treatments of ratings, rankings, and choices (Marshall and Bradlow 2002) to ordered-choice models with reporting heterogeneity (Greene and Hensher 2010; Falco et al. 2015) — supplies tools we adapt to the dual-response setting. To our knowledge, no prior work provides a utility-consistent joint model of a forced choice and an ordinal purchase evaluation.

The remainder of the paper proceeds as follows. Section 2 develops the behavioral model, derives the joint likelihood, establishes the nesting relationships, and presents the information-content and estimation results. Section 3 reports the simulation study. Section 4 presents the empirical application. Section 5 discusses implications for practice, including guidance on scale design, and Section 6 concludes.

2 Model

2.1 Setup

Consumer $i = 1, \ldots, N$ faces choice tasks $t = 1, \ldots, T$. Each task presents a set of "inside" goods $\mathcal{J}^+ = \{1, 2, \ldots, J\}$ alongside the ever-present "outside" good $j = 0$, so that the full choice set is $\mathcal{J} = \{0, 1, 2, \ldots, J\}$.[2] Consumer i derives utility u_{itj} from good j in task t:

$$u_{itj} = h(\mathbf{x}_{tj}, \boldsymbol{\beta}_i) + \eta_{itj}, \tag{1}$$

where $\mathbf{x}_{tj}$ is a length-P column vector of good characteristics, $\boldsymbol{\beta}_i$ is a length-P column vector of consumer-specific taste parameters, and η_{itj} encapsulates factors known to the consumer but unobserved by the researcher. We take $h(\mathbf{x}_{tj}, \boldsymbol{\beta}_i) = \mathbf{x}_{tj}'\boldsymbol{\beta}_i$ and write $V_{itj} \coloneqq \mathbf{x}_{tj}'\boldsymbol{\beta}_i$ for this deterministic component, but the linear specification is not required.

The outside good is special in two respects. First, it is associated with a zero vector of characteristics ($\mathbf{x}_{t0} = \mathbf{0}$), so that $V_{it0} = 0$; this is the standard normalization that anchors the location of utility (Haaijer et al. 2001). Second — and this is the central behavioral assumption of the paper — the consumer observes η_{itj} for all inside goods but *does not* observe η_{it0}. The

[2] We hold J fixed across tasks and consumers to lighten notation; nothing in what follows requires it.

inside goods are described completely on the screen in front of her; the outside good stands in for a purchase occasion that has not yet arrived, and her evaluation of it is uncertain.[3] [4]

In each task the consumer provides two responses:

1. **First response (choice).** She reports her preferred inside good,

$$j_{it}^{*} = \underset{j \in \mathcal{J}^{+}}{\operatorname{argmax}}\ u_{itj}. \tag{2}$$

   This is a *forced* choice: the outside good is not among the options.

2. **Second response (ordinal purchase evaluation).** She reports a value y_{it} on a discrete ordinal scale $w \in \mathcal{W} = \{1, \ldots, W\}$ (e.g., "very unlikely" to "very likely" on a five-point scale) that characterizes her evaluation of purchasing j_{it}^{*} — that is, of choosing j_{it}^{*} over the outside good. The binary dual-response format of Brazell et al. (2006) and Diener et al. (2006) is the special case $W = 2$.

We model $\eta_{itj} \overset{iid}{\sim} \text{Gumbel}\,(0, 1)$, independent of $\mathbf{x}_{tj}$ and independent across goods, tasks, and consumers. Two classical properties of the Gumbel family organize everything that follows (McFadden 1981; Cardell 1997).

Lemma 2.1. *Let* $u_{itj} = V_{itj} + \eta_{itj}$ *with* $\eta_{itj} \overset{iid}{\sim}$ *Gumbel*$(0, 1)$ *for* $j \in \mathcal{J}^{+}$*, and define* $u_{it}^{*} = \max_{j \in \mathcal{J}^{+}} u_{itj}$*. Then*

(i) $u_{it}^{*} \sim$ *Gumbel*$\,(\overline{\mu}_{it}, 1)$ *with location*

$$\overline{\mu}_{it} = \ln S_{it}, \qquad S_{it} \coloneqq \sum_{j \in \mathcal{J}^{+}} \exp\left(V_{itj}\right); \tag{3}$$

(ii) the identity *of the maximizer* j_{it}^{*} *and the* value *of the maximum* u_{it}^{*} *are statistically independent.*

Property (ii) is what makes the dual-response likelihood tractable: whatever the consumer tells us about the level of her best utility (the second response) is independent of which good delivered it (the first response), conditional on the deterministic utilities.

[3] This can be motivated by a framework in which $u_{it0} = 0$ and $u_{itj} = \mathbf{x}_{tj}{}'\boldsymbol{\beta}_i + \eta_{itj} - \eta_{it0}$ for $j \in \mathcal{J}^{+}$, where η_{it0} captures the consumer's uncertainty about her future tastes or circumstances at the moment of purchase. Because utilities are ordinal and η_{it0} is a common shock, it plays no role in the comparison among inside goods.

[4] While the parametrization $\eta_{it0} \sim \text{Gumbel}\,(0, 1)$ preserves symmetry among the $J + 1$ goods and is thus a natural choice, the framework easily accommodates alternative distributions for η_{it0}. For example, an affine function of individual characteristics could accommodate individual-level variation in the propensity to prefer the outside good.

2.2 First Response: Multinomial Logit

The first response is a standard forced choice among the inside goods. Under the Gumbel assumption,

$$\Pr\left(j_{it}^{*}=j\right)=\frac{\exp\left(V_{itj}\right)}{\sum_{k\in\mathcal{J}^{+}}\exp\left(V_{itk}\right)}=\frac{\exp\left(V_{itj}\right)}{S_{it}}, \tag{4}$$

the multinomial logit over inside goods only. Note that the outside good does not appear: the forced choice is uncontaminated by the consumer's uncertain evaluation of the outside good, which is precisely the design's appeal (Brazell et al. 2006).

2.3 Second Response: An Ordinal Model of the Purchase Evaluation

What does the consumer consult when answering "how likely are you to purchase the option you selected?" We develop two behavioral models that differ only in how the consumer treats the unobserved outside-good shock η_{it0}. Both lead to likelihoods of the same *cumulative-link* form, differing only in the link function, and both depend on the choice set only through the inclusive value $\overline{\mu}_{it}$.

2.3.1 Model A: Reporting a Purchase Probability

Under Model A the consumer takes her uncertainty about the outside good seriously and reports a *probability*. She knows u_{it}^{*} (she just identified j_{it}^{*}) but does not know η_{it0}, so from her perspective the probability that she would follow through and purchase j_{it}^{*} is

$$p_{it}=\Pr\left(\eta_{it0}<u_{it}^{*}\mid u_{it}^{*}\right)=F\left(u_{it}^{*}\right), \tag{5}$$

where $F(z)=\exp(-e^{-z})$ is the standard Gumbel CDF. The ordinal scale discretizes the unit interval: fix cut points $0=\alpha_0<\alpha_1<\cdots<\alpha_{W-1}<\alpha_W=1$ and suppose the consumer reports

$$y_{it}=w\quad\Longleftrightarrow\quad p_{it}\in\left[\alpha_{w-1},\alpha_{w}\right). \tag{6}$$

We assume consumers share the definitions of the qualitative scale labels, so the α_w are common across consumers (we relax this in Section 2.6.1); the interior cut points are unobserved by the researcher and are estimated.

Reporting $y_{it}=w$ is equivalent to reporting that $u_{it}^{*}\in\left[F^{-1}(\alpha_{w-1}),\,F^{-1}(\alpha_{w})\right)$, an interval statement about the consumer's best utility. From the researcher's perspective, $u_{it}^{*}\sim\text{Gumbel}(\overline{\mu}_{it},1)$ by Lemma 2.1(i), so

$$\begin{aligned}\Pr\left(y_{it}=w\right)&=F_{\text{Gumbel}(\overline{\mu}_{it},1)}\left(F^{-1}(\alpha_{w})\right)-F_{\text{Gumbel}(\overline{\mu}_{it},1)}\left(F^{-1}(\alpha_{w-1})\right)\\&=\alpha_{w}^{S_{it}}-\alpha_{w-1}^{S_{it}},\end{aligned} \tag{7}$$

using $F_{\text{Gumbel}(\overline{\mu},1)}(c) = \exp\left(-e^{-(c-\overline{\mu})}\right) = \exp\left(e^{\overline{\mu}} \ln \alpha\right) = \alpha^S$ when $c = F^{-1}(\alpha) = -\ln(-\ln \alpha)$. The second-response probabilities are differences of powers of the cut points, with the exponent $S_{it} = e^{\overline{\mu}_{it}}$ summarizing the attractiveness of the task's inside goods: the more attractive the set, the more probability mass shifts toward high scale points.

2.3.2 Model B: Grading a Resolved Comparison

Under Model B the consumer behaves as the binary dual-response literature implicitly assumes: she resolves (or acts as if she resolves) the outside-good shock and evaluates the realized utility difference

$$d_{it} = u^*_{it} - \eta_{it0}.$$

Rather than reporting only its sign — the binary "would you buy it?" — she grades its magnitude on the W-point scale: fix utility-scale cut points $-\infty = c_0 < c_1 < \cdots < c_{W-1} < c_W = \infty$ and suppose

$$y_{it} = w \quad \Longleftrightarrow \quad d_{it} \in [c_{w-1}, c_w)\,. \tag{8}$$

From the researcher's perspective, $u^*_{it} \sim \text{Gumbel}(\overline{\mu}_{it}, 1)$ and $\eta_{it0} \sim \text{Gumbel}(0, 1)$ are independent, so their difference is logistic: $d_{it} \sim \text{Logistic}\,(\overline{\mu}_{it}, 1)$. Hence

$$\Pr\left(y_{it} = w\right) = \Lambda\left(c_w - \overline{\mu}_{it}\right) - \Lambda\left(c_{w-1} - \overline{\mu}_{it}\right), \qquad \Lambda(z) = \frac{1}{1 + e^{-z}}. \tag{9}$$

This is an ordered logit (Greene and Hensher 2010) with the inclusive value $\overline{\mu}_{it}$ as the (negatively signed) linear predictor.

2.3.3 A Unified Cumulative-Link Representation

Define $c_w = F^{-1}(\alpha_w)$ in Model A so that both models carry utility-scale cut points. Then both second-response models take the cumulative-link form

$$\Pr\left(y_{it} \le w\right) = G\left(c_w - \overline{\mu}_{it}\right), \qquad G = \begin{cases} F \;\; \text{(standard Gumbel)} & \text{Model A,} \\ \Lambda \;\; \text{(standard logistic)} & \text{Model B.} \end{cases} \tag{10}$$

The behavioral story pins down the link: a consumer who *carries* her outside-good uncertainty into the report (Model A) generates a Gumbel — i.e., complementary log-log type — link, while a consumer who *resolves* it and grades the outcome (Model B) generates the familiar logistic link. The two links are numerically close over most of their range and diverge in the tails; which better describes respondents is an empirical question we take up in Section 3 and Section 4. Both are estimable at identical parameter counts ($W - 1$ interior cut points), so the comparison is a clean, non-nested horse race between links.

2.4 Joint Likelihood

By Lemma 2.1(ii), the value of u_{it}^* is independent of the identity of j_{it}^*; and η_{it0} is independent of the inside-good shocks. The two responses in a task are therefore independent conditional on $(\mathbf{x}_t, \boldsymbol{\beta}_i)$, and the per-task likelihood contribution is simply the product of Equation 4 and Equation 10:

$$\ell_{it}(\boldsymbol{\beta}_i, \mathbf{c}) = \underbrace{\frac{\exp(V_{itj^*})}{S_{it}}}_{\text{choice (MNL)}} \times \underbrace{\left[G(c_{y_{it}} - \overline{\mu}_{it}) - G(c_{y_{it}-1} - \overline{\mu}_{it})\right]}_{\text{ordinal purchase evaluation}}, \tag{11}$$

with $\mathbf{c} = (c_1, \ldots, c_{W-1})$ and the full likelihood $L_i = \prod_t \ell_{it}$. We emphasize that this factorization is not an *assumption* of independent errors across the two responses, as in the "DR-2Max" tradition (Diener et al. 2006; Sawtooth Software 2017); it is a *consequence* of the max-stability properties of the Gumbel family within a single, internally consistent utility draw. The same realized utilities generate both responses.

2.5 Special Cases and Relation to Existing Models

Binary dual response, Model B. Set $W = 2$ with the single cut point $c_1 = \gamma$. From Equation 9, $\Pr(\text{would buy}) = \Lambda(\overline{\mu}_{it} - \gamma) = S_{it}/(e^{\gamma} + S_{it})$, and the joint probabilities from Equation 11 telescope:

$$\Pr(j, \text{buy}) = \frac{\exp(V_{itj})}{e^{\gamma} + S_{it}}, \qquad \Pr(j, \text{not buy}) = \frac{\exp(V_{itj})}{S_{it}} \cdot \frac{e^{\gamma}}{e^{\gamma} + S_{it}}. \tag{12}$$

With $\gamma = 0$ these are *exactly* the joint probabilities of the "Unified Model" of Hung, Liu, et al. (2025) (their Eq. 4), which is in turn algebraically identical to the "DR-AnyMax" model of Diener et al. (2006); and Equation 12 with free γ is the standard single-stage MNL with a no-choice alternative-specific constant (Haaijer et al. 2001). Our ordinal model is thus a strict generalization of the leading binary dual-response model: at $W = 2$, Model B *is* that model, and for $W > 2$ the additional cut points are the only new parameters.

Binary dual response, Model A. Set $W = 2$ in Equation 7: $\Pr(\text{not buy}) = \alpha_1^{S_{it}} = F(c_1 - \overline{\mu}_{it})$, a *complementary log-log* rather than logistic binary response. Model A therefore implies that even the familiar binary dual response is misspecified — mildly, given the proximity of the links — when consumers genuinely face outside-good uncertainty at the moment they answer. This observation gives the binary special case of Model A independent interest: it provides a one-parameter alternative to the binary logit second stage that can be tested on existing binary dual-response data.

DR-2Max. The other binary model in the literature, "DR-2Max" (Diener et al. 2006), as implemented in Sawtooth Software's CBC/HB (Sawtooth Software 2017), models the second response as a binary logit on the *chosen good's* deterministic utility: $\Pr(\text{buy} \mid j^*) =$

$e^{V_{itj^*}} / \left(e^{V_{itj^*}} + e^{\gamma}\right)$. This requires drawing a fresh error for the chosen good in the second stage — the consumer who just declared j^* best re-evaluates it from scratch. Under our framework such conditioning is unnecessary: by Lemma 2.1(ii) the distribution of the best utility does not depend on which good attained it, so the second response should depend on the set only through $\overline{\mu}_{it}$. This distinction is testable, and the existing evidence favors the inclusive-value form: Diener et al. (2006) found DR-AnyMax fit dual-response data better than DR-2Max, and Hung, Liu, et al. (2025) reach the same conclusion for their Unified Model.

Practitioner heuristics for ordinal dual responses. When practitioners field an ordinal purchase-likelihood question today, common practice is either (i) *dichotomization* — collapse the scale at some point (e.g., top box or top two boxes) and estimate the binary model, discarding the ordinal information and hard-coding an arbitrary cut; or (ii) *constant-probability weighting* — map each scale point to a fixed purchase probability (e.g., "definitely would buy" = 0.9) independent of the choice set. Both are nested within, or dominated by, Equation 11: dichotomization is Equation 12 applied to a coarsened y_{it}, and constant-probability weighting violates Equation 7, under which the mapping from scale point to purchase probability necessarily depends on the attractiveness of the inside goods through S_{it}. These two heuristics are the benchmarks in our simulation study (Section 3).

2.6 Heterogeneity

We accommodate preference heterogeneity with a standard hierarchical Bayesian specification (Rossi et al. 2005; Train 2001):

$$\boldsymbol{\beta}_i \overset{iid}{\sim} \mathcal{N}\left(\overline{\boldsymbol{\beta}}, \Sigma_\beta\right), \tag{13}$$

with conjugate priors on $\left(\overline{\boldsymbol{\beta}}, \Sigma_\beta\right)$. In the main model the cut points $\mathbf{c}$ are common across consumers: the scale labels ("very likely," "somewhat likely," ...) are assumed to mean the same thing to everyone, while *preferences* vary.

2.6.1 Heterogeneous Cut Points

The assumption of a shared reporting scale is convenient but contestable: a rich literature documents individual differences in the use of ordinal response scales (Greene and Hensher 2010; Falco et al. 2015). We therefore develop an extension with consumer-specific cut points. Parameterize

$$c_{i1} \in \mathbb{R}, \qquad c_{iw} = c_{i,w-1} + \exp\left(\delta_{iw}\right) \text{ for } w = 2, \ldots, W-1, \tag{14}$$

and let $\boldsymbol{\theta}_i = (c_{i1}, \delta_{i2}, \ldots, \delta_{i,W-1})'$ follow a multivariate normal population distribution, estimated jointly with (and, if desired, correlated with) $\boldsymbol{\beta}_i$. The exponentiated increments enforce monotonicity without constrained sampling. Identification at the individual level is necessarily partial — with T tasks per consumer, individual cut points and individual preferences are

jointly informed by only T ordinal reports — so the hierarchical prior does real work here, shrinking individual scale usage toward the population norm. Section 3 examines both the payoff to modeling cut-point heterogeneity when it is present in the data-generating process and the robustness of the common-cut-point model when it is not.

2.7 Identification

Three normalizations do the identifying work. First, $\mathbf{x}_{t0} = \mathbf{0}$ anchors the location of utility: all inside-good utilities are measured relative to the (mean) outside good. Second, the unit Gumbel scale anchors the scale of utility, as in any logit model. Third, given these, the interior cut points are identified by the joint distribution of first and second responses across tasks that vary in attractiveness S_{it}: tasks with more attractive inside goods shift the ordinal response distribution upward at a rate governed by Equation 10, and the cut points are pinned down by the levels at which those shifts occur. One practical caution follows directly: the design matrix must not contain a constant common to all inside goods (attributes should be dummy- or effects-coded against reference levels), since a common constant shifts $\overline{\mu}_{it}$ one-for-one against the cut points. In the heterogeneous-cut-point extension, the population means of $\boldsymbol{\theta}_i$ are identified in the same way, while individual deviations are identified — and shrunk — through the hierarchy.

2.8 The Information Content of the Ordinal Response

How much does a W-point second response add over a binary one? The question has a clean answer in terms of Fisher information, extending the efficiency analysis of Hung, Liu, et al. (2025) from binary to ordinal second responses.

The per-task Fisher information about $\boldsymbol{\beta}$ decomposes into the familiar MNL term (Huber and Zwerina 1996) plus a second-response increment. Writing $\pi_w = G(c_w - \overline{\mu}) - G(c_{w-1} - \overline{\mu})$ and $g = G'$, the second response contributes

$$\mathcal{I}_{it}^{(2)} = \omega\left(\overline{\mu}_{it}; \mathbf{c}\right) \overline{\mathbf{x}}_{it}\, \overline{\mathbf{x}}_{it}', \qquad \omega = \sum_{w=1}^{W} \frac{\left[g\left(c_w - \overline{\mu}\right) - g\left(c_{w-1} - \overline{\mu}\right)\right]^2}{\pi_w}, \tag{15}$$

where $\overline{\mathbf{x}}_{it} = \sum_{j \in \mathcal{J}^+} \Pr\left(j_{it}^* = j\right) \mathbf{x}_{tj}$ is the choice-probability-weighted average attribute vector (the gradient of $\overline{\mu}_{it}$ with respect to $\boldsymbol{\beta}$). Two properties follow. First, the increment is positive semi-definite: the second response never destroys information, echoing the binary result of Hung, Liu, et al. (2025). Second, ω is non-decreasing under refinement of the scale partition — coarsening an ordinal response is a garbling in the sense of sufficiency, so the binary response ($W = 2$) is the *least* informative member of the family, and information rises toward the continuous-report limit as $W \to \infty$. The practical magnitudes — how much of the binary-to-continuous gap a five- or seven-point scale captures, and how the gains vary with task attractiveness — are quantified in Section 3.

2.9 Estimation

The aggregate (homogeneous-$\boldsymbol{\beta}$) model is estimated by maximum likelihood: the log of Equation 11 summed over consumers and tasks is smooth in $(\boldsymbol{\beta}, \mathbf{c})$ after the monotonicity-preserving reparameterization of the cut points, and standard quasi-Newton methods converge quickly.

The hierarchical model is estimated with a custom Markov chain Monte Carlo sampler, implemented in R with computational kernels in C++ (Rcpp), following the hybrid Gibbs architecture of Rossi et al. (2005): (1) a random-walk Metropolis step for each $\boldsymbol{\beta}_i$ (and $\boldsymbol{\theta}_i$ in the heterogeneous-cut-point extension) using the individual likelihood $\prod_t \ell_{it}$; (2) a random-walk Metropolis step for the common cut-point block on the unconstrained scale; and (3) conjugate draws for the population moments $\left(\overline{\boldsymbol{\beta}}, \Sigma_\beta\right)$. Model comparison uses the log marginal density approximated by the method of Newton and Raftery (1994), matching the practice in Hung, Liu, et al. (2025) and permitting direct comparisons among Model A, Model B, the binary dual-response models, and the practitioner heuristics. Complete replication code, together with three worked-example notebooks that build from the simplest homogeneous two-attribute case to the full model with heterogeneous cut points, accompanies the paper.

3 Simulation Study

The simulation study has four goals: (i) verify that the estimation routines recover the parameters of the data-generating process, for both the aggregate and hierarchical versions of the model; (ii) quantify what is lost by the two approaches practitioners currently apply to ordinal dual-response data — dichotomization and constant-probability weighting; (iii) probe robustness to the placement of the cut points, including data-generating processes in which respondents do *not* share a common interpretation of the scale; and (iv) quantify the information gains from finer ordinal scales, making the Fisher-information results of Section 2.8 concrete. Throughout, Model A (Section 2.3.1) and Model B (Section 2.3.2) are simulated and estimated in parallel, including cross-fit experiments in which data generated under one link are estimated under the other.

3.1 Design

The baseline design mirrors a typical commercial study. Each synthetic respondent completes $T = 12$ choice tasks; each task presents $J = 4$ inside goods described by $P = 8$ attribute levels (dummy-coded from a small factorial, plus a continuous price term); the ordinal follow-up uses a $W = 5$ point scale. Attribute levels are drawn to mimic a balanced fractional-factorial design. The true cut points place meaningful mass in every scale category at typical task attractiveness.

Three families of estimators are applied to every simulated dataset:

1. **The proposed model** (Equation 11), under the correct link and under the misspecified link (A estimated on B data and vice versa).
2. **Binary benchmarks:** the dual-response binary logit of Diener et al. (2006) and Hung, Liu, et al. (2025) applied to dichotomized data, with the scale collapsed at each possible cut point (top box, top two boxes, ...), so the best-case and worst-case dichotomizations are both visible.
3. **The constant-probability heuristic:** scale points mapped to fixed purchase probabilities. We implement two versions — equal-interval probabilities ($\Pr(y = w) = 1/W$ implied cut points) and cut points set to the observed marginal frequencies of the scale points — reflecting the variants encountered in practice.

Performance is evaluated on four criteria, matching the decisions the model supports: parameter recovery (bias and root mean squared error for $\boldsymbol{\beta}$, or for the individual-level $\boldsymbol{\beta}_i$ posteriors in the hierarchical case (Sarrias 2020)); in-sample model comparison via log marginal density (Newton and Raftery 1994); holdout prediction (hit rates for the first response, and predicted versus actual distributions of the ordinal response, on withheld tasks); and predicted purchase-incidence shares, the quantity practitioners ultimately report to clients.

3.2 Aggregate Model

We first simulate with homogeneous preferences ($\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i) and estimate by maximum likelihood.

3.3 Hierarchical Model

We then simulate with $\boldsymbol{\beta}_i \sim \mathcal{N}(\overline{\boldsymbol{\beta}}, \Sigma_\beta)$ and estimate the hierarchical Bayesian model with the MCMC sampler of Section 2.9.

3.4 The Cost of Current Practice

This subsection collects the comparisons that answer the practitioner's question: *how much does the correct model actually matter?*

3.5 Cut-Point Placement and Scale-Usage Heterogeneity

Two robustness experiments target the cut points. First, *extreme placements*: data-generating processes in which one or more cut points sit far into the tails (e.g., a "very likely" category that is almost never used), which stress the estimation of the associated α_w or c_w and mimic scales with too many points for the construct being measured. Second, *heterogeneous scale usage*: data generated with respondent-specific cut points $\boldsymbol{\theta}_i$ as in Section 2.6.1, estimated (a) with the common-cut-point model, to measure the robustness of preference estimates to ignoring reporting heterogeneity, and (b) with the heterogeneous-cut-point model, to measure what the richer specification buys and how much data it needs.

3.6 Information Gains from Finer Scales

Finally, we evaluate Equation 15 numerically across $W \in \{2, 3, 5, 7, 11\}$, holding the design fixed and placing cut points at population quantiles, to trace how the second response's Fisher-information increment grows with scale refinement — the ordinal extension of the design-efficiency results in Brazell et al. (2006) and Hung, Liu, et al. (2025). We report both the analytic D-efficiency gains and their finite-sample counterparts (RMSE ratios from the Monte Carlo replications), and translate them into the practitioner's currency: the reduction in respondents (or tasks) required to match the accuracy of a binary dual-response study.

4 Empirical Application

We illustrate the model with data from a commercial choice-based conjoint study conducted in partnership with a consumer insights consultancy, in which each choice task was followed by an ordinal purchase-likelihood question.

4.1 Data

4.2 Model Specifications and Estimation

We estimate six specifications on the same data: Model A and Model B (Section 2), each in aggregate and hierarchical form; the binary dual-response benchmark applied to the dichotomized scale (top-two-box); and the constant-probability heuristic. Hierarchical specifications use the sampler of Section 2.9.

4.3 Results

4.4 Scale-Usage Heterogeneity

5 Discussion

5.1 What the Behavioral Model Changes

The premise that the consumer observes everything about the inside goods but not the outside-good shock does more than motivate a likelihood; it reorganizes how the dual-response literature's models relate to one another. In the binary literature, the choice between DR-2Max and DR-AnyMax was framed as a question about the second question's wording — whether the respondent evaluates her selected option or the whole set (Diener et al. 2006) — and the shared-error Unified Model of Hung, Liu, et al. (2025) was motivated by the internal consistency of a single utility draw. Our framework arrives at the inclusive-value form from the primitive assumption about information: because the maximum of Gumbel utilities is independent of which good attains it, *any* coherent report about the best option's standing against the outside good must depend on the task only through its inclusive value. The empirical success of the AnyMax/Unified form over DR-2Max, documented in both Diener et al. (2006) and Hung, Liu, et al. (2025), is what this framework predicts. Where our account differs from Brazell et al. (2006) is in what the second response *is*: not a second choice from an augmented set, but a report about an unresolved comparison — which is what licenses asking for it on an ordinal scale in the first place.

The distinction between Model A and Model B is, likewise, a substantive behavioral question dressed as a link function. Model A takes the respondent's uncertainty literally — she reports a probability — while Model B assumes she resolves the outside-good draw and grades the outcome. Their binary special cases already separate the two: Model B recovers the standard logit second stage, while Model A implies a complementary log-log second stage even for yes/no data. Because the two links diverge mainly in the tails, discriminating between them requires either large samples or designs with wide variation in task attractiveness; the model comparison in Section 4 provides evidence from one commercial dataset, but we regard the question as open and worth testing on the many existing binary dual-response datasets, where the Model A binary case can be estimated with a one-line change to current code.

5.2 Implications for Practice

Three points bear directly on applied work. First, if an ordinal dual response has been fielded, dichotomizing it is costly and arbitrary: the simulation results in Section 3.4 show that the analyst's choice of where to cut the scale moves managerially relevant quantities, while the model uses every scale point and removes the choice. Second, constant-probability weighting has a specific, diagnosable flaw: the purchase probability that a scale point implies is not a constant

— under Equation 7 it depends on the attractiveness of the displayed set through S_{it} — so calibrations struck on one design will not transfer to another. Third, the information results in Section 2.8 and Section 3.6 give concrete design guidance: the follow-up question's scale should be chosen the way sample sizes are chosen, by information budgeting. Our results suggest most of the binary-to-continuous information gap is captured by a five- to seven-point scale, after which respondent burden and sparsely used categories argue against finer partitions.

5.3 Relation to the Outside-Good Literature

Our model holds the outside good's *mean* utility fixed at zero and locates all action in the consumer's uncertainty about its realization. Two adjacent strands suggest natural extensions. Hung, Liu, et al. (2025) show that a budget constraint applied to the second response — but not the forced choice — improves fit and materially changes equilibrium price predictions; the same asymmetric treatment is available here, censoring the ordinal report for options priced above a respondent-level budget, and the ordinal response should sharpen the identification of that budget since near-threshold options will earn lukewarm rather than merely negative reports. Hung, Kurz, et al. (2025) document that the no-choice option's utility can drift over the course of a survey as respondents learn about the category; in our notation this is drift in the cut points (or in the outside-good location), and the task-order diagnostics of Section 4.1 carry over directly. More broadly, the opt-out literature's warning that a single constant rarely captures heterogeneous non-purchase behavior (Campbell and Erdem 2019; Haaijer et al. 2001) applies to the dual response as well; the heterogeneous-cut-point extension of Section 2.6.1 is one response, and richer structures (e.g., serial non-purchasers as a latent class) are compatible with the framework.

5.4 Limitations

Four limitations qualify the results. First, the Gumbel error assumption is load-bearing: the factorization of the joint likelihood and the inclusive-value form of the second response both rest on max-stability, and relaxing the error family (e.g., toward mixed or nested structures within a task) would break the closed form, though not the framework — the second response would simply require simulation over the max distribution. Second, we treat tasks as independent draws, abstracting from learning, fatigue, and reference formation across the sequence (Hung, Kurz, et al. 2025). Third, the common-cut-point model assumes a shared reading of the scale labels; the extension of Section 2.6.1 relaxes this, but individual-level identification is necessarily thin with commercial task counts, and our recommendation to treat scale-usage heterogeneity as a robustness check rather than a default reflects that. Fourth, like all stated-preference methods, the model calibrates stated rather than revealed purchase behavior; the ordinal report inherits whatever hypothetical bias attends purchase-intent questions generally (Morrison 1979), and linking the estimated cut points to realized purchase rates — the ordinal analogue of the calibration literature on binary intentions — is a natural next step.

6 Conclusion

Dual-response conjoint designs separate a question consumers find easy — which of these do you like best? — from one they find genuinely uncertain — would you actually buy it? This paper takes that uncertainty seriously as a modeling primitive. When the consumer observes everything about the displayed alternatives but not her valuation of the outside good, the natural follow-up question is ordinal, and the natural model is the one derived here: a multinomial logit for the forced choice joined to a cumulative-link ordinal model on the task's inclusive value, with the link function determined by whether the respondent reports her purchase probability directly (Model A) or grades a resolved comparison (Model B). The framework nests the leading binary dual-response models as its two-point special case, so nothing is lost relative to current best practice; a W-point scale strictly adds Fisher information, and the common practitioner shortcuts — dichotomizing the scale or assigning fixed purchase probabilities to its points — are shown to be, respectively, wasteful and internally inconsistent. Simulation evidence confirms that the aggregate and hierarchical Bayesian estimators recover the model's parameters, including the cut points that give the scale labels quantitative meaning, and the empirical application demonstrates the model on commercial data.

Beyond the specific model, the paper's broader point is that response formats and statistical models should be developed together. The ordinal purchase-likelihood question earned its place in practice because it matches how respondents experience purchase decisions; it lacked only a likelihood. Supplying one turns an informal fielding habit into a measurement instrument — and opens a research agenda that includes calibrating estimated cut points against realized purchase behavior, exploiting the ordinal response to identify budget constraints, and optimal design of the scale itself.

References

Allenby, Greg M., Nino Hardt, and Peter E. Rossi. 2019. "Economic Foundations of Conjoint Analysis." In *Handbook of the Economics of Marketing*, edited by Jean-Pierre Dubé and Peter E. Rossi, vol. 1. North-Holland.

Brazell, Jeff D., Christopher G. Diener, Ekaterina Karniouchina, William L. Moore, Válerie Séverin, and Pierre-Francois Uldry. 2006. "The No-Choice Option and Dual Response Choice Designs." *Marketing Letters* 17 (4): 255–68. https://doi.org/10.1007/s11002-006-7943-8.

Campbell, Danny, and Seda Erdem. 2019. "Including Opt-Out Options in Discrete Choice Experiments: Issues to Consider." *The Patient - Patient-Centered Outcomes Research* 12 (1): 1–14. https://doi.org/10.1007/s40271-018-0324-6.

Cardell, N. Scott. 1997. "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity." *Econometric Theory* 13 (2): 185–213. https://doi.org/10.1017/s0266466600005727.

Dhar, Ravi, and Itamar Simonson. 2003. "The Effect of Forced Choice on Choice." *Journal of Marketing Research* 40 (2): 146–60. https://doi.org/10.1509/jmkr.40.2.146.19229.

Diener, Christopher G., Bryan K. Orme, and David Yardley. 2006. "Dual Response 'None' Approaches: Theory and Practice." *Proceedings of the Sawtooth Software Conference* (Sequim, WA), 157–67.

Falco, Paolo, William F. Maloney, Bob Rijkers, and Mauricio Sarrias. 2015. "Heterogeneity in Subjective Wellbeing: An Application to Occupational Allocation in Africa." *Journal of Economic Behavior & Organization* 111: 137–53. https://doi.org/10.1016/j.jebo.2014.12.022.

Greene, William H., and David A. Hensher. 2010. "Ordered Choices and Heterogeneity in Attribute Processing." *Journal of Transport Economics and Policy* 44 (3): 331–64.

Haaijer, Rinus, Wagner A. Kamakura, and Michel Wedel. 2001. "The 'No-Choice' Alternative in Conjoint Choice Experiments." *International Journal of Market Research* 43 (1): 93–106. https://doi.org/10.1177/147078530104300105.

Huber, Joel, and Klaus Zwerina. 1996. "The Importance of Utility Balance in Efficient Choice Designs." *Journal of Marketing Research* 33 (3): 307–17. https://doi.org/10.1177/002224379603300305.

Hung, Chien-Yu, Peter Kurz, Roger A. Bailey, Joel Huber, and Greg M. Allenby. 2025. "Re-Examining the No-Choice Option in Conjoint Analysis." *Journal of Choice Modelling* 57: 100578. https://doi.org/10.1016/j.jocm.2025.100578.

Hung, Chien-Yu, Yueming Liu, Jeff D. Brazell, and Greg M. Allenby. 2025. "Dual Response in Conjoint Analysis." *Marketing Letters*, ahead of print. https://doi.org/10.1007/s11002-025-09787-1.

Juster, F. Thomas. 1966. "Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design." *Journal of the American Statistical Association* 61 (315): 658–96. https://doi.org/10.1080/01621459.1966.10480897.

Marshall, Pablo, and Eric T. Bradlow. 2002. "A Unified Approach to Conjoint Analysis Models." *Journal of the American Statistical Association* 97 (459): 674–82. https://doi.org/10.1198/016214502388618410.

McFadden, Daniel L. 1981. *Structural Discrete Probability Models Derived from Theories of Choice*. The MIT Press.

Morrison, Donald G. 1979. "Purchase Intentions and Purchase Behavior." *Journal of Marketing* 43 (2): 65–74. https://doi.org/10.1177/002224297904300207.

Newton, Michael A., and Adrian E. Raftery. 1994. "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap." *Journal of the Royal Statistical Society: Series B (Methodological)* 56 (1): 3–48. https://doi.org/10.1111/j.2517-6161.1994.tb01956.x.

Rossi, Peter E., Greg M. Allenby, and Robert McCulloch. 2005. *Bayesian Statistics and Marketing*. John Wiley & Sons.

Sarrias, Mauricio. 2020. "Individual-Specific Posterior Distributions from Mixed Logit Models: Properties, Limitations and Diagnostic Checks." *Journal of Choice Modelling* 36: 100224. https://doi.org/10.1016/j.jocm.2020.100224.

Sawtooth Software. 2017. *The CBC System for Choice-Based Conjoint Analysis, Version 9*. Technical Paper. Sawtooth Software, Inc.

Schlereth, Christian, and Bernd Skiera. 2017. "Two New Features in Discrete Choice Experiments to Improve Willingness-to-Pay Estimation That Result in SDR and SADR: Separated (Adaptive) Dual Response." *Management Science* 63 (3): 829–42. https://doi.org/10.1287/mnsc.2015.2367.

Train, Kenneth E. 2001. *A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit*. Working Paper. University of California, Berkeley.