

Imperfect Attention in Discrete Choice Experiments

Sen Zeng (Cornell University)

Discussion by

Jing Li
CHOICE Institute
University of Washington

Annual Health Econometrics Workshop
Hunter College, CUNY
October 2, 2026

Motivation: Do Respondents Process What DCEs Vary?

- Discrete Choice Experiments (DCE) are useful in studying policies that have not happened yet (e.g., a menthol ban) or when observational data is not available
- Identification assumes **every respondent** processes **every varied attribute**
- A muted response to an attribute is ambiguous: weak taste, or inattention? Choice data alone cannot tell (Caplin 2016)
- This paper:** combine choices with post-task recall to separate **tastes** from **attribute attention**
- Compares a **full-attention mixed logit** with parametric and neural-network **attention priors** on fit, prediction, and policy counterfactuals

	Non-Menthol Cigarettes	Menthol Cigarettes	Tobacco Flavored E-cigarettes	Menthol Flavored E-cigarettes	None
Product					
Price	\$10.00	\$5.00	\$8.00	\$4.00	
Legality	Legal	Legal	Legal	Prohibited. Strictly enforced. Only available from illegal dealers, e.g. street sellers	I will quit smoking cigarettes and not use e-cigarettes

Example choice task from the menthol DCE (Fig. 2)

Suppose your doctor offers the following options. **Which would you choose?**

	Treatment A	Treatment B	Neither
Chance symptoms improve	60 in 100	75 in 100	I would not take either treatment
Risk of serious side effect	1 in 100	5 in 100	
How it is taken	Pill, once daily	Injection, once a month	
Time to feel benefit	2 weeks	3 months	
Clinic visits per year	4	1	
Travel time to clinic	15 minutes	1 hour	
Out-of-pocket cost per month	\$50	\$10	
	☐ Treatment A	☐ Treatment B	☐ Neither

Application: A Menthol Prohibition DCE (Kenkel et al. 2025)

639 menthol smokers; 12 tasks each; 5 alternatives; 2 attributes (price, legality)

Alternative	Price	Legality / availability
Non-menthol cigarettes	$0.5p, p, 2p$	Legal
Menthol cigarettes	$0.5p, p, 2p$	Legal, retail-illegal, street-illegal
Tobacco-flavored e-cigarettes	\$2, \$4, \$8	Legal
Menthol-flavored e-cigarettes	\$2, \$4, \$8	Legal, retail-illegal, street-illegal
Quit smoking and using e-cigarettes	—	—

- p = price paid for the last pack
- **Retail-illegal:** sold under the counter and online
- **Street-illegal:** only from illegal dealers

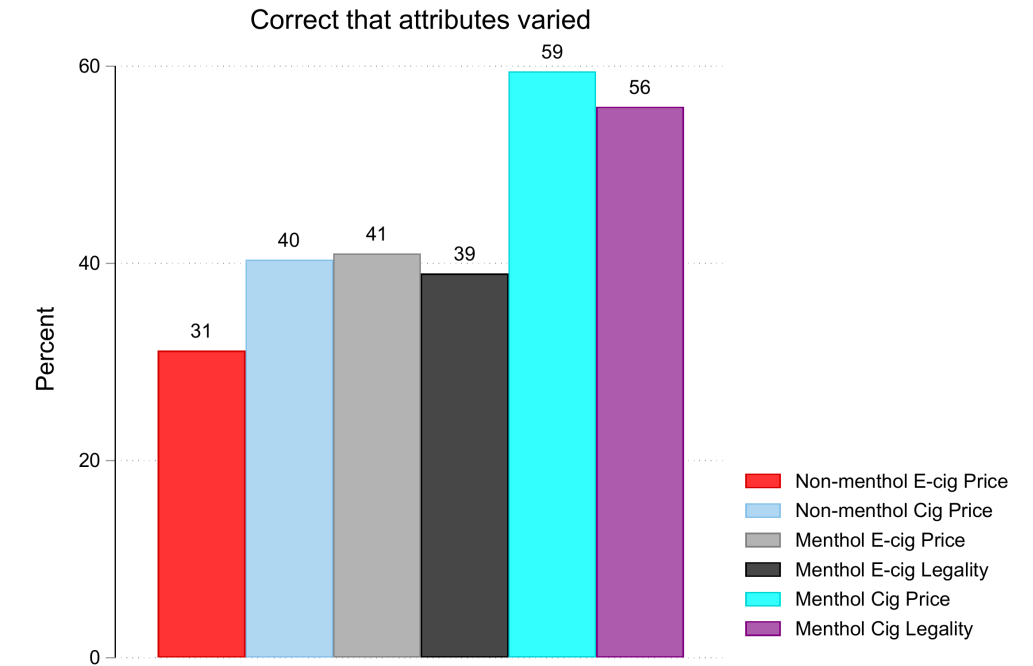
Descriptive Evidence for Imperfect Attention: Inertia and Weak Recall

Choice patterns

	Used in past year	Share of choices	Same option in all tasks
Menthol cigarettes	100.00%	43.21%	12.05%
Non-menthol cigarettes	—	8.93%	0.16%
Menthol e-cigarettes	41.56%	17.75%	2.03%
Tobacco-flavored e-cigarettes	18.92%	8.01%	0.31%
Quit	60.67%	22.10%	4.54%

- **About 19%** choose the same option in all 12 tasks
- Could be attachment, limited consideration, or skipped attributes

Recall of product-attribute variation



- **Only 31–59% correct recall**
- Highest for menthol cigarettes

Model: Mixed Logit with Latent Attribute Attention

$$U_{ijt} = \alpha_{ij} + \beta_{ip} A_{ij}^P p_{ijt} + A_{ij}^L (\delta_{ij}^R D_{ijt}^R + \delta_{ij}^S D_{ijt}^S) + \varepsilon_{ijt}$$

Utility

- U_{ijt} : utility of respondent i for alternative j in task t
- α_{ij} : alternative-specific constants, normal (quit is the base)
- p_{ijt} : price; β_{ip} : price coefficient, negative lognormal
- D_{ijt}^R, D_{ijt}^S : retail- and street-illegal indicators for menthol products; $\delta_{ij}^R, \delta_{ij}^S$: negative lognormal
- ε_{ijt} : type-I extreme value

Attention

- $A_{ij}^P, A_{ij}^L \in \{0, 1\}$: attends to the price or legality of alternative j
- Six dimensions (four prices, two menthol-legality attributes): $2^6 = 64$ respondent-level states, fixed across all 12 tasks
- If a respondent does not pay attention to an attribute, that attribute drops out of her utility
- **Benchmark**: every attention indicator set to one

Choices and Recall Jointly Identify Attention

$$\tilde{L}_i(\theta, \psi, \eta) = \sum_{a \in \{0,1\}^6} \underbrace{p_\psi(a \mid W_i)}_{\text{attention prior}} \underbrace{p_\eta(R_i \mid a)^\omega}_{\text{recall measurement}} \underbrace{\int p_\theta(y_i \mid a, \beta) f_\theta(\beta) d\beta}_{\text{panel choice likelihood}}$$

Attention prior

- **W: DCE completion time, phone use, multiple visits (attention shifter)**
- Parametric logit or one-layer neural network

Recall measurement

- Correct / ambiguous / none, for 6 attribute dimensions
- Noisy auxiliary measurement of attention
- Monotonicity: recall more likely if attentive
- **Recall down-weighted: $\omega = 0.1$**

Choice likelihood

- Mixed logit given attention state
- 200 Halton draws

- Estimated jointly by generalized expectation-maximization

What the Model Assumes

Separating tastes from attention rests on 5 assumptions/restrictions:

1. Attention is a vector of binary variables and fixed at the respondent level across all 12 tasks
2. Attention dimensions are independent given W , with extra mass on full attention
3. W shifts attention but has no direct effect on utility (exclusion restriction)
4. Each recall response depends only on attention to that attribute, not on tastes or choices
5. Tastes are independent of attention

What the Model Assumes

Separating tastes from attention rests on five assumptions/restrictions:

1. Attention is a vector of binary variables and fixed at the respondent level across all 12 tasks
2. Attention dimensions are independent given W , with extra mass on full attention
- 3. W shifts attention but has no direct effect on utility (exclusion restriction)**
4. Each recall response depends only on attention to that attribute, not on tastes or choices
- 5. Tastes are independent of attention**

Results

Mean coefficients

Mean coefficient	Benchmark	Parametric	ML
Menthol cigarette constant	4.472	5.608	5.627
Price (\$)	-0.383	-0.439	-0.440
Menthol cigarettes: retail-illegal	-1.543	-3.015	-3.055
Menthol cigarettes: street-illegal	-2.149	-4.205	-4.229
Menthol e-cigarettes: retail-illegal	-1.533	-3.076	-3.070
Menthol e-cigarettes: street-illegal	-2.536	-4.690	-4.703

In-sample fit

	Benchmark	Parametric attention	ML attention
Choice log-likelihood	-6,736	-6,607	-6,588
AIC	13,509	13,276	13,269
BIC	13,589	13,414	13,479

- **Legality coefficients roughly double** once attention is modeled
- Price mean more negative ($-0.38 \rightarrow -0.44$); price SD falls ($0.65 \rightarrow 0.48$)
- Log-likelihood improves by 129–148 points; Vuong tests favor both attention models

Prediction Performance

New respondents (5-fold)

	Held-out pseudo- R^2 Higher is better	Share RMSE Lower is better
Benchmark	0.1527	0.0243
Parametric attention	0.1452	0.0344
ML attention	0.1449	0.0347

- **Benchmark wins** when predicting new respondents from the prior alone

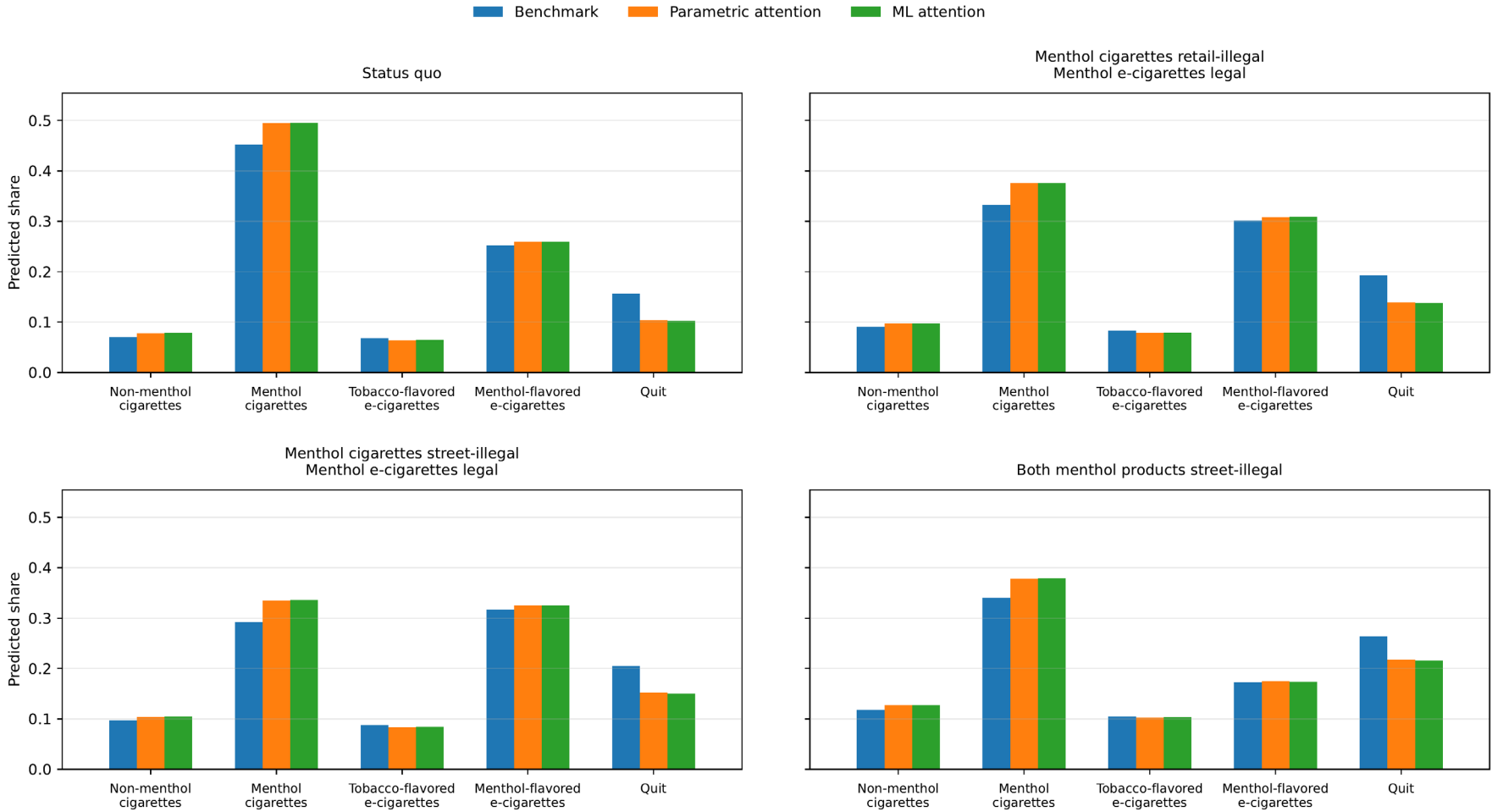
Tasks 7–12, after observing tasks 1–6

	Held-out pseudo- R^2 Higher is better	Share RMSE Lower is better
Benchmark	0.4619	0.0111
Parametric attention	0.4957	0.0123
ML attention	0.4955	0.0125

- **Attention models achieve better fit** on individual choices once early choices are observed

The paper reports the prediction performance openly and interprets the model as structural, not a better forecaster

Counterfactuals



Quit share: benchmark
→ parametric

15.7% → 10.4%

Status quo

26.4% → 21.8%

Both menthol products street-illegal

+10.7 → +11.4

Percentage-point change from
own baseline

Discussion: What This Paper Does Well

- Tackles an important and under-appreciated problem for applied DCE users
- Uses practical auxiliary data: a recall block any online survey can add
- Keeps utility structural and confines machine learning to the attention prior
- Reports honestly, including where the model does not win
- Concise and clearly written

Model Assumptions: Recap

Separating tastes from attention rests on 5 assumptions/restrictions:

1. Attention is a vector of binary variables and fixed at the respondent level across all 12 tasks
2. Attention dimensions are independent given W , with extra mass on full attention
3. **W shifts attention but has no direct effect on utility (exclusion restriction)**
4. Each recall response depends only on attention to that attribute, not on tastes or choices
5. **Tastes are independent of attention**

Identification: Tastes Imputed from Attentive Respondents

The legality taste is identified only from respondents classified as attentive, then applied to those classified as inattentive

- Valid only under Assumption #5 (tastes independent of attention), which is built into the likelihood but not explicitly discussed
- Rational inattention, which motivates the paper, predicts the opposite: people who care little about an attribute rationally ignore it (see also Collins, Rose & Hensher 2013)
- If so, the doubled legality coefficients may overstate the disutility of inattentive respondents

Suggestion: Allow ASCs and price tastes to differ by legality-attention state, and let W shift the legality taste; report whether tastes vary with attention.

Identification: Is W Excluded from Utility?

Assumption #3 requires W (survey-process measures) to shift attention but not tastes; but completion time etc. plausibly reflects both

- W = DCE completion time, phone use, multiple survey visits
- **Reverse causality:** a committed menthol smoker may answer quickly because the choice is easy. Strong tastes could cause short completion time
- **Confounding:** phone use and repeat visits correlate with age, income, and functional status, which plausibly shift tastes
- Minor: give concrete examples of W where it is first introduced (§3.4)

Suggestion: let W enter taste means as a check; in future designs, randomized attribute order or position could be attention shifters excluded from utility by design

Identification: How Much Work Does Recall Do?

Recall data are the paper's key addition to a standard attribute non-attendance model, yet they enter with one-tenth weight ($\omega = 0.1$)

- **As $\omega \rightarrow 0$** , attention is identified from choices and functional form alone, as in Hole (2011)
- The value 0.1 is motivated only informally (recall is noisy), yet it determines how much the results rest on recall versus functional form

Suggestion: results for $\omega \in \{0.05, 0.1, 0.5, 1\}$; report recall probabilities by attention state; consider choosing ω by cross-validation

Interpretation: Survey Attention vs. a Real Ban

The counterfactuals assume that attention under a real ban equals attention in the survey: a natural default, but not explicitly stated

- A respondent who skimmed past "Prohibited" in the survey is assumed to overlook the ban in real life
- A real ban is salient: news coverage, empty shelves, retailer signs
- The quit-share gap between models is already present at the status quo; the change from each model's own baseline is similar

Quit share

	Benchmark	Parametric	ML
Status quo quit share	15.66%	10.41%	10.29%
Both menthol products street-illegal	26.38%	21.76%	21.61%
Change from own status quo	+10.71 pp	+11.35 pp	+11.32 pp

Suggestion: state the assumption explicitly; show counterfactuals with legality attention set to 1 as a bound

Practice: Better Attention Data

The recall measures are coarse: one post-task measure per attribute, so attention must be fixed across all 12 tasks

- **Task-level timing:** most survey platforms record time per page; if available, attention could vary by task and capture fatigue or learning
- **Designed attention shifters:** randomize attribute order or on-screen position across respondents; because it is random, it is unrelated to tastes and can serve as W in the current respondent-level model, addressing #3

Practice: When Does Modeling Attention Matter?

The model nests full attention, so it can serve as a diagnostic: in a well-designed DCE it should match the benchmark

- This particular DCE application is demanding: 5 alternatives including potentially unfamiliar products, and policy attributes that respondents may have little prior related experience in
- Simplifying and improving DCE design could go a long way in ensuring better attention
- Highly useful for applied researchers doing DCE to know whether their DCE is one where attention changes the answer

Suggestion (future research): a simulation varying inattention rates and design complexity, showing when the attention model and the benchmark diverge; turn it into practical guidance on when to add a recall block

Summary

The paper tackles an important problem, offers a practical solution, and reports results honestly.

My main comments/suggestions:

1. Identification

- Test whether tastes vary with attention; let W shift tastes
- Show sensitivity to ω

2. Interpretation

- State the attention assumption behind the counterfactuals; bound with legality attention = 1

3. Practice

- Use or recommend task-level timing and designed attention shifters
- Simulate when modeling attention changes the answer