WEB APPENDIX

The Customer Journey as a Source of Information

These materials have been supplied by the authors to aid in the understanding of their paper.

A Specification of model components

The model is informed by four types of behaviors: queries ($\mathbf{q}_{ij}$), clicks (y^c_{ijt}), filters ($\mathbf{f}_{ij}$), and purchases (y^p_{ij}).

A.1 Query

We leverage the information in the search query through multiple query variables that capture the context of a journey. These variables help capture journey-specific needs, even for different journeys of the same customer. For example, day-of-the-week and time-of-the-day may be relevant to infer whether a search query in a food delivery platform relates to a single-person weekday lunch vs. a romantic Friday night dinner. In a travel setting, the number of passengers, dates of travel, and destination, among others, can be informative about the context for what the user might be looking for.

We denote by $\mathbf{q}_{ij}$ the vector of query variables that describe journey j for customer i,

$$\mathbf{q}_{ij} = \begin{bmatrix} q_{ij1} & \dots & q_{ijM} \end{bmatrix}',$$

where each component indexed by $m \in \{1, \dots, M\}$ describes a different type of query variable (e.g., length of the stay, traveling with kids). Because these pieces of information are provided by the customer to obtain a set of product results that match their preferences, we treat each query variable as an outcome that depends on some unobserved component that captures the customer's true need in the focal journey.[1] We model $\mathbf{q}_{ij}$ as a function of a vector of parameters $\boldsymbol{\omega}_j = \begin{bmatrix} \omega_{j1} & \dots & \omega_{jM} \end{bmatrix}'$.

We assume that given $\boldsymbol{\omega}_j$, the components of $\mathbf{q}_{ij}$ are conditionally independent, that is:

$$p(\mathbf{q}_{ij} | \boldsymbol{\omega}_j) = \prod_{m=1}^{M} p(q_{ijm} | \omega_{jm}). \tag{1}$$

[1] Potentially, customers could slightly modify the query along the journey while searching for a product to satisfy the same need (e.g., changing the departing date when customers search for flight tickets). We model only the first query by a customer in each journey due to the minimal additional information these often provide. That being said, the model can easily be extended to learn from multiple query instances.

Each type of query variable m could be of multiple types: (1) binary, (2) categorical, (3) continuous real-valued, or (4) continuous positive-valued. We flexibly model q_{ijm} using a different distribution $p_m(q_{ijm}|\omega_{jm})$ for each type of variable m,

$$q_{ijm} \sim \begin{cases} \text{Bernoulli}(\omega_{jm}) & \text{if } q_{ijm} \text{ is binary} \\ \text{Categorical}(\omega_{jm}) & \text{if } q_{ijm} \text{ is categorical} \\ \exp(\omega_{jm}) & \text{if } q_{ijm} \text{ is continuous positive-valued} \\ \mathcal{N}(\omega_{jm}, \sigma_m^2) & \text{if } q_{ijm} \text{ is continuous,} \end{cases} \tag{2}$$

where each parameter ω_{jm} has the appropriate support given the distribution it governs.[2] Our model can accommodate other distributions such as Poisson or Binomial for count variables, and Student's t-distribution or Cauchy for long-tailed continuous variables.

A.2 Joint model of clicks and purchase

We structure the modeling of clicks and purchase decisions in two phases. First, customers explore products through clicks and potential filtering to form a consideration set. Next, customers proceed to the purchase decision stage, where they either choose an item from their considered set or decide not to make a purchase. All of these decisions are guided by a shared set of customer preferences, denoted as $\boldsymbol{\beta}_{ij}$.

Click decisions Along the journey, the customer clicks through pages of product results. The customer can navigate back and forth between clicking on products and refining their searches. In each step, the customer decides among: (1) clicking on one of the products shown on the page to consider it for purchase, (2) continuing to search to receive a new set of results (e.g., by adjusting the query or filtering the results), or (3) ending the search and moving to the purchase decision among those considered.

We model the click decision of alternative k at step t of the journey using a discrete choice model. We define Page_{ijt} as the set of products displayed to customer i in journey j at step t. The customer faces a decision between: clicking on one of the displayed products $k \in \text{Page}_{ijt}$, continue searching to get a new set of products ($k = s$), or finish the search process and move to the purchase decision ($k = e$), which could mean either purchasing a considered product or deciding not to buy. We denote the choice consumer i makes at step t of journey j by $y^c_{ijt} \in \text{Page}_{ijt} \cup \{s, e\}$,

[2] We choose to define σ_m fixed across all journeys, to avoid the issue of singularity. That is analogous to approaches that prevent regularity issues commonly found when estimating Gaussian mixtures with component-specific variances (Bishop 2006). These issues emerge when the mean of one of the Gaussian components is equal to a single data point, which leads to a term contributing to the model likelihood that grows to infinity as the variance of such component goes to zero.

which we model using a multinomial probit specification with latent propensities u^c_{ijtk}, such that

$$y^c_{ijt} = \underset{k \in \text{Page}_{ijt} \cup \{s,e\}}{\arg\max} \left\{u^c_{ijtk}\right\}, \text{ with}$$

$$u^c_{ijtk} = \begin{cases} \beta^{0c}_{ij} + \mathbf{x}^c_{ijtk}{}' \cdot \boldsymbol{\beta}^x_{ij} + \text{log-rank}_{ijtk} \cdot \eta + \varepsilon_{ijtk} & \text{if } k \in \text{Page}_{ijt}, \\ \beta^{0s}_{ij} + \varepsilon_{ijts} & \text{if } k = s, \\ \beta^{0e}_{ij} + \varepsilon_{ijte} & \text{if } k = e, \end{cases} \tag{3}$$

where $\varepsilon_{ijtk} - \varepsilon_{ijte} \sim \mathcal{N}(0, \sigma^2)$, $\mathbf{x}^c_{ijtk}$ is the vector of attributes of product k, $\boldsymbol{\beta}^x_{ij}$ is the vector of customer- and journey-specific product-attribute preferences, β^{0c}_{ij} is the intercept for clicking on a product, β^s_{ij} is the intercept for the decision to continue searching, and β^{0e}_{ij} is the intercept for finishing the search process, normalized to 0 for identification purposes. Note that by definition, customers stop searching in the last observed step and move to the purchase decision (i.e., $y^c_{ijT_{ij}} = e$).

We control for ranking effects on search (Ursu 2018) by incorporating the log of the position of product k within the results page into the search in u^c_{ijtk} and using η to capture such ranking effects.[3] Such a term also captures search costs within a page, along with the intercepts in (3) that capture users' propensity to keep searching and are related to search costs across pages. We denote the vector of product-attributes $\mathbf{x}^c_{ijtk}$ to be t-specific to allow for a subset of all attributes $\mathbf{x}_{ijk}$ to be shown differently in different types of pages. For example, while customers observe all attributes at the moment of purchase, they may not observe all of them on certain pages while searching. Similarly, the observability of certain attributes may even differ among different types of pages (e.g., departing and returning results pages for flights in online travel). For example, the attributes of the return leg of a flight are not shown when customers choose the first leg of the flight, so they drop from the choice model.

Filter decisions Websites and apps usually collect other types of interactions — e.g., whether a user filters results based on some attributes — information that can be used to further inform about customer preferences in that particular journey. Unlike clicks, filters are not frequently observed in the data — many journeys do not have filters, and when they do, they generally occur only once along the entire journey. We avoid computational burden by modeling the filtering decision at the overall journey level (rather than at the step level t); that is, we model whether the customer uses a particular filter *at any time* during the journey.

We denote by $\ell \in \{1, \ldots, L_{ij}\}$ the level customer i in journey j can filter on and define $\mathbf{f}_{ij}$ to be the vector of summarized filter decisions for customer i in journey j,

$$\mathbf{f}_{ij} = \begin{bmatrix} f_{ij1} & \ldots & f_{ijL_{ij}} \end{bmatrix}',$$

[3]Following Ursu (2018), we include the log-position rank in the click decision but not in the purchase-given-clicks decision.

where each component $f_{ij\ell}$ is defined by

$$f_{ij\ell} = \begin{cases} 1 & \text{if customer } i \text{ filters on level } \ell \text{ within journey } j \\ 0 & \text{otherwise.} \end{cases}$$

We model each component $\mathbf{f}_{ij}$ using a binary probit specification such that

$$f_{ij\ell} \sim \text{Bernoulli}\left(\Phi\left(\alpha_\ell^0 + \mathbf{w}_{ij\ell}{}' \cdot \boldsymbol{\alpha}_\ell^w + \boldsymbol{\beta}_{ij}^x{}' \cdot \boldsymbol{\alpha}_\ell^\beta\right)\right), \tag{4}$$

where α_ℓ^0 is the intercept of filtering on level ℓ, $\boldsymbol{\beta}_{ij}^x$ is the same set of preferences that drive clicks and purchases, and $\boldsymbol{\alpha}_\ell^\beta$ is the vector that relates those preferences to the filtering decision. It is this term that allows the model to learn preferences for attributes by fusing filtering decisions about those attributes. To control for other factors that may affect the filtering decisions (e.g., the overall characteristics of unfiltered results), we include $\mathbf{w}_{ij\ell}$ capturing a set of controls that summarize the set of (unfiltered) results. In particular, we control in $\mathbf{w}_{ij\ell}$ for the number of total products, the percentage of products with level ℓ, and the number of top 5 products with level ℓ in the unfiltered results.[4]

Purchase given consideration After clicking and possibly filtering through product results, customers make the purchase decision. We model this as a discrete choice among the alternatives in a consideration set $\mathcal{C}_{ij}$. Specifically, we define the consideration set as the set of products that have been clicked on at least once during the course of the journey, plus the outside option of not purchasing

$$\mathcal{C}_{ij} = \left\{k \ : \ k \in \text{Page}_{ijt},\ y_{ijt}^c = k,\ t \in \{1, \ldots, T_{ij}\}\right\}. \tag{5}$$

We model the purchase decision using a multinomial probit specification with latent propensities u_{ijk}^p. That is,

$$\begin{aligned} y_{ij}^p &= \underset{k \in \mathcal{C}_{ij} \cup \{\text{NoPurchase}\}}{\arg\max} \left\{u_{ijk}^p\right\}, \text{ where} \\ u_{ijk}^p &= \begin{cases} \beta_{ij}^{0p} + \mathbf{x}_{ijk}{}' \cdot \boldsymbol{\beta}_{ij}^x + \epsilon_{ijk} & \text{if } k \in \mathcal{C}_{ij} \\ \beta_{ij}^{0o} + \epsilon_{ijo} & \text{if } k = \text{NoPurchase}, \end{cases} \end{aligned} \tag{6}$$

with $\epsilon_{ijk} - \epsilon_{ijo} \sim \mathcal{N}(0, \sigma_p^2)$, and where $\mathbf{x}_{ijk}$ is the vector of attributes of product k, $\boldsymbol{\beta}_{ij}^x$ is the same vector of customer- and journey-specific product-attribute preferences from (3), β_{ij}^{0p} is the intercept for purchasing a product, and β_{ij}^{0o} is the intercept for not buying, normalized to 0 for identification purposes.

[4]As journeys may contain multiple unfiltered results due to multi-session journeys, we average these controls across the unfiltered results pages of all sessions within a journey.

Finally, we define

$$\boldsymbol{\beta}_{ij} = \left(\beta^{0c}_{ij},\ \beta^{0s}_{ij},\ \beta^{0p}_{ij},\ \boldsymbol{\beta}^{x\prime}_{ij}\right)', \tag{7}$$

as the vector of all clicks and purchase preferences.

B Model priors

B.1 Distributions

We detail the specification of the prior distribution for the model parameters.

First, for the population covariance matrix Σ that governs customer heterogeneity in (7), we choose the standard Wishart prior for the precision matrix Σ^{-1},

$$\Sigma^{-1} \sim \text{Wishart}(r_0, R_0).$$

Second, we put priors on the Pitman-Yor process discount and strength parameters, d and a,[5] respectively by

$$\begin{aligned} d &\sim \text{Beta}(\phi^d_0, \phi^d_1) \\ a &\sim \text{Gamma}(\phi^a_0, \phi^a_1). \end{aligned}$$

Third, we put priors on the location parameters θ_c by defining the base distribution of the Pitman-Yor process, F_0. As described in (11), the location parameters are drawn from $\theta_c \sim F_0(\boldsymbol{\phi}_0)$. Following the notation in (8), consider θ^ω and θ^ρ as the components of θ that correspond to query parameters $\boldsymbol{\omega}_j$ and click-purchase parameters $\boldsymbol{\rho}_j$, respectively. We define F_0 as a multivariate distribution factorized by each of the components of θ, defined by

$$F_0(\theta|\boldsymbol{\phi}_0) = \left(\prod_{m=1}^{M} F^\omega_{0m}(\theta^\omega_m|\phi_{0m})\right) \times \mathcal{N}(\theta^\rho|\mu_0, V_0),$$

where we assume Gaussian priors for the location parameter of click and purchase preferences, and F^q_{0m} is defined accordingly to the support of the parameter that governs the distribution of each query variable m described in (2). That is,

$$F^\omega_{0m}(\theta^\omega_m|\phi_{0m}) = \begin{cases} \text{Beta}(\phi_{0ma}, \phi_{0mb}) & \text{if } q_{ijm} \text{ is binary} \\ \text{Dirichlet}(\phi_{0m}) & \text{if } q_{ijm} \text{ is categorical} \\ \text{Gamma}(\phi_{0ma}, \phi_{0mb}) & \text{if } q_{ijm} \text{ is continuous positive-valued} \\ \mathcal{N}(\phi_{0m\mu}, \phi_{0m\sigma}) & \text{if } q_{ijm} \text{ is continuous.} \end{cases}$$

[5]We restrict the model to $a > 0$.

Finally, we put mean-zero Gaussian priors on all other parameters in the model including η in (3), and α^0_ℓ, $\boldsymbol{\alpha}^w_\ell$ and $\boldsymbol{\alpha}^\beta_\ell$ in (4)

$$\begin{aligned}
\eta &\sim \mathcal{N}(0, s^2_\eta) \\
\boldsymbol{\alpha}^0 &\sim \mathcal{N}(\mathbf{0}, S_{\alpha,0}) \\
\boldsymbol{\alpha}^w_\ell &\sim \mathcal{N}(\mathbf{0}, S_{\alpha,w}),\ \forall \ell \\
\boldsymbol{\alpha}^\beta_\ell &\sim \mathcal{N}(\mathbf{0}, S_{\alpha,\beta}),\ \forall \ell
\end{aligned}$$

We use the following hyperparameters for the model priors. For the Wishart priors, we use $r_0 = n_\beta + 5$ and $R_0 = \frac{1}{r_0 - n_\beta + 1} \cdot \mathbf{I}_{n_\beta}$, where n_β is the dimensionality of $\boldsymbol{\beta}_{ij}$. For the Pitman-Yor discount and strength priors we use $\phi^d_0 = 0.5$, $\phi^d_1 = 5.0$, $\phi^a_0 = 1.0$, and $\phi^a_1 = 0.1$. For the Pitman-Yor base distribution we use the following hyperparameters: (a) $\phi_{0ma} = \phi_{0mb} = 2$ for Beta; (b) $\phi_{0m} = \mathbf{1}$ for Dirichlet, (c) $\phi_{0ma} = \phi_{0mb} = 2$ for Gamma; $\phi_{0m\mu} = 0$ and $\phi^2_{0m\sigma} = 100$ for Gaussian; and (d) $\mu_0 = 0$ and $V_0^{-1} = 10^4$ for θ^ρ. For the position effect priors, we use $s^2_\eta = 25$. Finally, for the priors governing the filter component, we use $S^{-1}_{\alpha,0} = \frac{1}{25} \cdot \mathbf{I}$, $S^{-1}_{\alpha,\omega} = \frac{1}{25} \cdot \mathbf{I}$, and $S^{-1}_{\alpha,\beta} = 2.5 \cdot \mathbf{I}$.

C Blocked-Gibbs sampler algorithm

Our Metropolis-within-Gibbs MCMC sampling algorithm is based on Ishwaran and James (2001) approximation using the stick-breaking representation of the Pitman-Yor (PY) Process, truncating the infinite mixture by setting $V_C = 1$ for a large enough integer C. This approximation allows us to draw context memberships of different journeys in parallel, significantly increasing our sampling scheme's speed. We use adaptive Metropolis-Hastings (M-H) steps to update the PY parameters d and a as these full conditionals do not have a closed form (a has closed form only if $d = 0$). We use Gibbs steps for all other parameters as their full conditionals have closed form. Similarly to the click and purchase components, we use data augmentation for the filter decisions and define $u^f_{ij\ell} = \alpha^0_\ell + \mathbf{w}_{ij\ell}' \cdot \boldsymbol{\alpha}^w_\ell + \boldsymbol{\beta}^x_{ij}{}' \cdot \boldsymbol{\alpha}^\beta_\ell + \varepsilon^f_{ij\ell}$, such that $\varepsilon^f_{ij\ell} \sim \mathcal{N}(0,1)$ and $f_{ij\ell} = \mathbb{1}(u^f_{ij\ell} > 0)$.

We sequentially update the parameters by,

1. Draw latent click utilities for alternative $k \in \text{Page}_{ijt} \cup \{s\}$ using a truncated Gaussian by,

$$u^c_{ijtk} \sim \begin{cases} \text{Truncated-}\mathcal{N}\left(\bar{u}^c_{ijtk}, 1, \text{lower} = -\infty, \text{upper} = 0\right) & \text{if } y^c_{ijt} = e \\ \text{Truncated-}\mathcal{N}\left(\bar{u}^c_{ijtk}, 1, \text{lower} = \max\{u^c_{ijt-k}, 0\}, \text{upper} = \infty\right) & \text{if } y^p_{ij} = k \\ \text{Truncated-}\mathcal{N}\left(\bar{u}^c_{ijtk}, 1, \text{lower} = -\infty, \text{upper} = \max\{u^c_{ijt-k}\}\right) & \text{otherwise,} \end{cases}$$

where $\bar{u}^c_{ijtk} = \beta^{0c}_{ij} + \mathbf{x}^c_{ijtk}{}' \cdot \boldsymbol{\beta}^x_{ij} + \text{log-rank}_{ijtk} \cdot \eta$ if $k \in \text{Page}_{ijt}$, and $\bar{u}^c_{ijtk} = \beta^{0s}_{ij}$ if $k = s$.

2. Draw latent purchase utilities by,

$$u_{ijk}^{p} \sim \begin{cases} \text{Truncated-}\mathcal{N}\left(\bar{u}_{ijk}^{p}, 1, \text{lower} = -\infty, \text{upper} = 0\right) & \text{if } y_{ij}^{p} = \text{NoPurchase} \\ \text{Truncated-}\mathcal{N}\left(\bar{u}_{ijk}^{p}, 1, \text{lower} = \max\{u_{ij-k}^{p}, 0\}, \text{upper} = \infty\right) & \text{if } y_{ij}^{p} = k \\ \text{Truncated-}\mathcal{N}\left(\bar{u}_{ijk}^{p}, 1, \text{lower} = -\infty, \text{upper} = \max\{u_{ij-k}^{p}\}\right) & \text{otherwise,} \end{cases}$$

where $\bar{u}_{ijk}^{p} = \beta_{ij}^{0p} + \mathbf{x}_{ijk}{}' \cdot \boldsymbol{\beta}_{ij}^{x}$.

3. Draw latent filter utilities by,

$$u_{ij\ell}^{f} \sim \begin{cases} \text{Truncated-}\mathcal{N}\left(\alpha_{\ell}^{0} + \mathbf{w}_{ij\ell}{}' \cdot \boldsymbol{\alpha}_{\ell}^{w} + \boldsymbol{\beta}_{ij}^{x}{}' \cdot \boldsymbol{\alpha}_{\ell}^{\beta}, 1, \text{lower} = -\infty, \text{upper} = 0\right) & \text{if } f_{ij\ell} = 0 \\ \text{Truncated-}\mathcal{N}\left(\alpha_{\ell}^{0} + \mathbf{w}_{ij\ell}{}' \cdot \boldsymbol{\alpha}_{\ell}^{w} + \boldsymbol{\beta}_{ij}^{x}{}' \cdot \boldsymbol{\alpha}_{\ell}^{\beta}, 1, \text{lower} = 0, \text{upper} = \infty\right) & \text{if } f_{ij\ell} = 1. \end{cases}$$

4. Draw individual-level stable preferences $\boldsymbol{\mu}_i$. We define a vector of click, purchase, and filter latent utilities for journey j,

$$\tilde{u}_{ij} = \begin{bmatrix} \left[u_{ijtk}^{c} - \text{log-rank}_{ijtk} \cdot \eta\right]_{tk} \\ \left[u_{ijk}^{p}\right]_{k} \\ \left[u_{ij\ell}^{f} - \alpha_{\ell}^{0} - \mathbf{w}_{ij\ell}{}' \cdot \boldsymbol{\alpha}_{\ell}^{w}\right]_{\ell} \end{bmatrix}^{\top},$$

and $\widetilde{\mathbf{X}}_{ij}$ the corresponding "stacked" matrix of vectors multiplying $\boldsymbol{\beta}_{ij}$ in equations (3), (4), and (6). That is,

$$\widetilde{\mathbf{X}}_{ij} = \begin{bmatrix} \widetilde{\mathbf{X}}_{ij}^{c} \\ \widetilde{\mathbf{X}}_{ij}^{p} \\ \widetilde{\mathbf{A}}_{ij}^{f} \end{bmatrix},$$

where $\widetilde{\mathbf{X}}_{ij}^{c}$ is the matrix of stacked click covariates. Specifically,

$$\widetilde{\mathbf{X}}_{ij}^{c} = \begin{bmatrix} \widetilde{\mathbf{X}}_{ij1}^{c} \\ \vdots \\ \widetilde{\mathbf{X}}_{ijt}^{c} \\ \vdots \\ \widetilde{\mathbf{X}}_{ijT_{ij}}^{c} \end{bmatrix} \quad \text{and} \quad \widetilde{\mathbf{X}}_{ijt}^{c} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & \mathbf{x}_{ijt1}^{c}{}' \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \mathbf{x}_{ijtk}^{c}{}' \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \mathbf{x}_{ijtK_{ijt}}^{c}{}' \end{bmatrix}.$$

Similarly, $\widetilde{\mathbf{X}}_{ij}^{p}$ is the matrix of stacked purchased covariates,

$$\widetilde{\mathbf{X}}_{ij}^{p} = \begin{bmatrix} 0 & 0 & 1 & \mathbf{x}_{ij1}{}' \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & \mathbf{x}_{ijk}{}' \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & \mathbf{x}_{ijtK_{ij}}{}', \end{bmatrix},$$

and $\widetilde{\mathbf{A}}_{ij}^{f} = \left[\boldsymbol{\alpha}_1^{\beta} \ldots \boldsymbol{\alpha}_{\ell}^{\beta} \ldots \boldsymbol{\alpha}_{L_{ij}}^{\beta}\right]'$.

The columns of each of these matrices multiply $\boldsymbol{\beta}_{ij} = \left(\beta_{ij}^{0c},\ \beta_{ij}^{0s},\ \beta_{ij}^{0p},\ \boldsymbol{\beta}_{ij}^{x}{}'\right)'$, respectively; which yields the terms in (3), (18), and (6).

We further define $\widetilde{\mathbf{X}}_i$ as

$$\widetilde{\mathbf{X}}_i = \begin{bmatrix} \widetilde{\mathbf{X}}_{i1} \\ \vdots \\ \widetilde{\mathbf{X}}_{ij} \\ \vdots \\ \widetilde{\mathbf{X}}_{iJ_i} \end{bmatrix},$$

and $\widetilde{\mathbf{u}}_i$ as

$$\widetilde{\mathbf{u}}_i = \begin{bmatrix} \widetilde{u}_{i1} - \widetilde{\mathbf{X}}_{i1} \cdot \boldsymbol{\rho}_1 \\ \vdots \\ \widetilde{u}_{ij} - \widetilde{\mathbf{X}}_{ij} \cdot \boldsymbol{\rho}_j \\ \vdots \\ \widetilde{u}_{iJ_i} - \widetilde{\mathbf{X}}_{iJ_i} \cdot \boldsymbol{\rho}_{J_i} \end{bmatrix}.$$

Finally, we draw $\boldsymbol{\mu}_i \sim \mathcal{N}(\widetilde{\mu}_i, \widetilde{S}_i)$ where

$$\begin{aligned} \widetilde{S}_i^{-1} &= \Sigma^{-1} + \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i \\ \widetilde{\mu}_i &= \widetilde{S}_i\left(\Sigma^{-1} \cdot \mathbf{0} + \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{u}}_i\right). \end{aligned}$$

5. Draw context membership z_j as follows

$$p(z_j = c|\cdot) = \frac{\pi_c \mathcal{P}_{jc}}{\sum\limits_{c'=1}^{C} \pi_{c'} \mathcal{P}_{jc'}},$$

where $\mathcal{P}_{jc} = \left(\prod_{m=1}^{M} p(q_{ijm}|\theta^{\omega}_{cm})\right) \cdot p\left(\tilde{u}_{ij} - \widetilde{\mathbf{X}}_{ij}\boldsymbol{\mu}_i|\widetilde{\mathbf{X}}_{ij}\theta^{\rho}_j, 1\right)$, with $p(q_{ijm}|\theta^{\omega}_{cm})$ denoting the pdf of query variables as defined in (2), and $p\left(\tilde{u}_{ij} - \widetilde{\mathbf{X}}_{ij} \cdot \boldsymbol{\mu}_i|\widetilde{\mathbf{X}}_{ij} \cdot \theta^{\rho}_j, 1\right)$ denoting the product of elementwise normal pdf evaluated at each components of $\tilde{u}_{ij} - \widetilde{\mathbf{X}}_{ij} \cdot \boldsymbol{\mu}_i$ with mean $\widetilde{\mathbf{X}}_{ij} \cdot \theta^{\rho}_j$ and variance 1.

6. Draw the query components of context location parameters θ^{ω}_c for each context c. We denote $\mathcal{J}(c)$ as the set of journeys j and $n_c = |\mathcal{J}(c)|$ as the number of journeys such that $z_j = c$. For each query variable m, we draw θ^{ω}_{cm} depending on the type of query variable modeled in (2). Specifically,

$$\theta^{\omega}_{cm} \sim \begin{cases} \text{Beta}\left(\phi_{0ma} + \sum\limits_{j \in \mathcal{J}(c)} q_{ijm}, \phi_{0mb} + n_c - \sum_{j \in \mathcal{J}(c)} q_{ijm}\right) & \text{if } q_{ijm} \text{ is binary} \\ \text{Dirichlet}\left(\phi_{0m} + [nq_{cm1}, \ldots, nq_{cmN_m}]^{\top}\right) & \text{if } q_{ijm} \text{ is categorical} \\ \text{Gamma}\left(\phi_{0ma} + n_c, \phi_{0mb} + \sum\limits_{j \in \mathcal{J}(c)} q_{ijm}\right) & \text{if } q_{ijm} \text{ is continuous positive-valued} \\ \mathcal{N}\left(\tilde{\mu}_{cm}, \tilde{s}_{cm}\right) & \text{if } q_{ijm} \text{ is continuous,} \end{cases}$$

where $nq_{cmn} = \sum_{j \in \mathcal{J}(c)} \mathbb{1}(q_{ijm} = n)$, $\tilde{s}^{-1}_{cm} = \left[\phi^{-1}_{0m\sigma} + \sigma^{-2}_m\right]$ and $\tilde{\mu}_{cm} = \tilde{s}_{cm} \sum_{j \in \mathcal{J}(c)} q_{ijm}$.

7. Draw the click-purchase context location parameters θ^{ρ}. We define $\bar{\mathbf{X}}_c$ and $\bar{\mathbf{u}}_c$ as

$$\bar{\mathbf{X}}_c = \left[\left[\widetilde{\mathbf{X}}_{i(j)j}\right]_{j \in \mathcal{J}(c)}\right], \text{ and } \qquad \bar{\mathbf{u}}_c = \left[\left[\tilde{u}_{i(j)j} - \widetilde{\mathbf{X}}_{i(j)j} \cdot \boldsymbol{\mu}_{i(j)}\right]_{j \in \mathcal{J}(c)}\right],$$

where $i(j)$ denotes the customer to whom journey j belongs to.

We draw $\theta^{\rho}_c \sim \mathcal{N}(\bar{\mu}_c, \bar{S}_c)$, where

$$\begin{aligned} \bar{S}^{-1}_c &= V^{-1}_0 + \bar{\mathbf{X}}'_c \bar{\mathbf{X}}_c \\ \bar{\mu}_c &= \bar{S}_c \left(V^{-1}_0 \cdot \mu_0 + \bar{\mathbf{X}}'_c \bar{\mathbf{u}}_c\right). \end{aligned}$$

8. Draw ranking effect η. Defining $\mathbf{r}$ as the vector of all log-rank$_{ijtk}$ values, and the vector of differences in click utilities $\mathbf{u}^{\mathrm{r}} = \left[\{u^c_{ijtk} - \beta^{0c}_{ij} + \mathbf{x}^{c}_{ijtk}{}' \cdot \boldsymbol{\beta}^x_{ij}\}_{ijtk}\right]$, we draw η by

$$\eta \sim \mathcal{N}(\bar{\mu}_{\eta}, \bar{s}^2_{\eta}),$$

where

$$
\begin{aligned}
\overline{s_\eta}^{-1} &= s_\eta^{-1} + \mathbf{r}'\mathbf{r} \\
\bar{\mu}_\eta &= \bar{s}_\eta \left(s_\eta^{-1} \cdot 0 + \mathbf{r}'\mathbf{u}^{\mathrm{r}} \right).
\end{aligned}
$$

9. Draw $\boldsymbol{\alpha}^0$. We define $\tilde{\mathbf{u}}_0^f$ as the vector of residual filter utilities where each component is an observation (i, j, ℓ) defined as

$$
\tilde{u}_{0ij\ell}^f = u_{ij\ell}^f - \mathbf{w}_{ij\ell}{}' \cdot \boldsymbol{\alpha}_\ell^w - \boldsymbol{\beta}_{ij}^x{}' \cdot \boldsymbol{\alpha}_\ell^\beta.
$$

We also define a binary matrix that multiplies the vector of intercepts $\boldsymbol{\alpha}^0$ to yield the respective level for each observation. In other words, this matrix encodes in binary variables the level ℓ to which the observation (row) belongs, such that the entry in each row that represents the observation (i, j, ℓ) takes the value one for column ℓ, and zero for all others. Consequently, we draw $\boldsymbol{\alpha}^0$ by

$$
\boldsymbol{\alpha}^0 \sim \mathcal{N}(\bar{\boldsymbol{\mu}}_{\alpha,0}, \bar{S}_{\alpha,0}),
$$

where

$$
\begin{aligned}
\bar{S}_{\alpha,0}^{-1} &= S_{\alpha,0}^{-1} + \tilde{\mathbf{w}}_0'\tilde{\mathbf{w}}_0 \\
\bar{\boldsymbol{\mu}}_{\alpha,0} &= \bar{S}_{\alpha,0} \left(S_{\alpha,0}^{-1} \cdot 0 + \tilde{\mathbf{w}}_0'\tilde{\mathbf{u}}_0^f \right).
\end{aligned}
$$

10. Draw $\boldsymbol{\alpha}_\ell^w$. We define $\tilde{\mathbf{u}}_{w,\ell}^f = \left[\{u_{ij\ell}^f - \alpha_\ell^0 - \boldsymbol{\beta}_{ij}^x{}' \cdot \boldsymbol{\alpha}_\ell^\beta\}_{ij} \right]$ as the vector of residual filter utilities, and $\mathbf{W}_\ell = \left[\{\mathbf{w}_{ij\ell}'\}_{ij} \right]$ the matrix of filter controls for level ℓ, and draw $\boldsymbol{\alpha}_\ell^w$ by

$$
\boldsymbol{\alpha}_\ell^w \sim \mathcal{N}(\bar{\boldsymbol{\mu}}_\ell^w, \bar{S}_{\alpha,w,\ell}),
$$

where

$$
\begin{aligned}
\bar{S}_{\alpha,w,\ell}^{-1} &= S_{\alpha,w}^{-1} + \mathbf{W}_\ell'\mathbf{W}_\ell \\
\bar{\boldsymbol{\mu}}_\ell^w &= \bar{S}_{\alpha,w,\ell} \left(S_{\alpha,w}^{-1} \cdot 0 + \mathbf{W}_\ell'\tilde{\mathbf{u}}_{w,\ell}^f \right).
\end{aligned}
$$

11. Draw $\boldsymbol{\alpha}_\ell^\beta$. We define $\tilde{\mathbf{u}}_{w,\ell}^b = \left[\{u_{ij\ell}^f - \alpha_\ell^0 - \mathbf{w}_{ij\ell}' \cdot \boldsymbol{\alpha}_\ell^w\}_{ij} \right]$ as the vector of residual filter utilities, and $\mathbf{B}_\ell = \left[\{\boldsymbol{\beta}_{ij}'\}_{ij} \right]$ as the matrix of preferences, where each row of the matrix contains the vector of preferences corresponding to the respective row in $\tilde{\mathbf{u}}_{w,\ell}^b$. Draw $\boldsymbol{\alpha}_\ell^\beta$ by

$$
\boldsymbol{\alpha}_\ell^\beta \sim \mathcal{N}(\bar{\boldsymbol{\mu}}_\ell^\beta, \bar{S}_{\alpha,\beta,\ell}),
$$

where

$$\bar{S}^{-1}_{\alpha,\beta,\ell} = S^{-1}_{\alpha,\beta} + \mathbf{B}'_\ell \mathbf{B}_\ell$$
$$\bar{\boldsymbol{\mu}}^{\beta}_{\ell} = \bar{S}_{\alpha,\beta,\ell}\left(S^{-1}_{\alpha,\beta} \cdot 0 + \mathbf{B}'_\ell \tilde{\mathbf{u}}^{b}_{w,\ell}\right).$$

12. (M-H step) Draw a proposal $a^{\text{prop}} \sim p_{a-\text{prop}}(\cdot|a)$. Update $a = a^{\text{prop}}$ with probability

$$\alpha(a, a^{\text{prop}}) = \min\left\{1,\ \frac{\text{Gamma}\left(a^{\text{prop}}|\phi^a_0, \phi^a_1\right)}{\text{Gamma}\left(a|\phi^a_0, \phi^a_1\right)} \cdot \frac{\prod_{c=1}^{C-1} \text{Beta}\left(V_c|1-d\ ,\ a^{\text{prop}} + c \cdot d\right)}{\prod_{c=1}^{C-1} \text{Beta}\left(V_c|1-d\ ,\ a + c \cdot d\right)} \cdot \frac{p_{a-\text{prop}}(a|a^{\text{prop}})}{p_{a-\text{prop}}(a^{\text{prop}}|a)}\right\}.$$

We use a log-normal $p_{a-\text{prop}}(\cdot|a) = \log\mathcal{N}(\log(a),\ \tau^2_n)$, where we use a vanishing adaptation procedure (Atchadé and Rosenthal 2005) to adapt the proposal step size to target an acceptance rate of 0.44 (Gelman et al. 1995) through

$$\tau^2_n = \begin{cases} \tau^2_0 & n \leqslant 200, \\ |\tau^2_{n-1} + \frac{\epsilon}{n}(ap_n - 0.44)| & n > 200, \end{cases}$$

where ap_n is the empirical acceptance rate up to iteration n. Note the proposal distribution is not symmetric and yields a ratio $\frac{p_{a-\text{prop}}(a|a^{\text{prop}})}{p_{a-\text{prop}}(a^{\text{prop}}|a)} = \frac{a^{\text{prop}}}{a}$.

13. (M-H step) Draw a proposal $d^{\text{prop}} \sim p_{d-\text{prop}}(\cdot|d)$. Update $d = d^{\text{prop}}$ with probability

$$\alpha(d, d^{\text{prop}}) = \min\left\{1,\ \frac{\text{Beta}\left(d^{\text{prop}}|\phi^d_0, \phi^d_1\right)}{\text{Beta}\left(d|\phi^d_0, \phi^d_1\right)} \cdot \frac{\prod_{c=1}^{C-1} \text{Beta}\left(V_c|1-d^{\text{prop}}\ ,\ a + c \cdot d^{\text{prop}}\right)}{\prod_{c=1}^{C-1} \text{Beta}\left(V_c|1-d\ ,\ a + c \cdot d\right)} \cdot \frac{1/p_{d-\text{prop}}(d^{\text{prop}}|d)}{1/p_{d-\text{prop}}(d|d^{\text{prop}})}\right\}.$$

We use a logit-normal proposal distribution $p_{d-\text{prop}}(\cdot|d) = \text{logit}-\mathcal{N}(\text{logit}(d),\ s^2_n)$, where the logit function is defined by $\text{logit}(d) = \log\left(\frac{d}{1-d}\right)$, and the logit-normal pdf is defined by $\text{logit}-\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x(1-x)} \exp\left\{-\frac{(\text{logit}(x)-\mu)^2}{2\sigma^2}\right\}$. We adapt s^2_n analogously to τ^2_n in the previous step.

14. Draw context probabilities π_c, by drawing the stick parameters V_c from

$$V_c \sim \text{Beta}\left(1 - d + n_c\ ,\ a + c \cdot d + \sum_{c'=c+1}^{C} n_{c'}\right),$$

and compute π_c according to (12).

15. Draw population covariance matrix Σ, by

$$\Sigma^{-1} \sim \text{Wishart}(r_1, R_1),$$

where

$$r_1 = r_0 + I$$
$$R_1^{-1} = R_0^{-1} + \sum_i \boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_i'$$

D Posterior distribution of holdout journey preferences

We outline the procedure to update the posterior distribution of preferences for holdout journeys, given data on the focal journey and past journeys. (This corresponds to the right-hand side of (14)). There are several relevant considerations for this procedure.

First, we leverage the Pitman-Yor process when making inferences on new journeys, by allowing for a previously unobserved context to be discovered in this focal journey. Second, as the posterior of global parameters is obtained with a large number of journeys in the training sample, we approximate the posterior of these parameters given all training data plus focal journey j, by the posterior without focal journey j. That is, the inference on global parameters remains largely unchanged by the addition of a single journey (except for discovering a context that has not been observed before, as mentioned above). This assumption allows us to maintain computational efficiency by not re-estimating the whole model when updating the inference on current journey preferences as new data arrives. Third, as commented in Section 2.4, the context of past journeys is conditionally dependent on the focal journey given past and current journey data, because stable preferences and contexts both jointly determine the outcomes in both journeys. Therefore, in the process of drawing preferences for new journeys, we update the inferences for past journeys of the focal customer as well.

For each customer i, we denote the focal (holdout) journey by j, with j' referring to journeys different from the focal one. The set of past journeys (not including j) is denoted by $\mathcal{J}(i)$, the vector of contexts of all past journeys by $\mathbf{z}_{i,-j} = \{z_{ij'}\}_{j' \in \mathcal{J}(i)}$, the entire journey data for a journey j' by $\mathcal{H}_{i,j'} = \{\mathbf{q}_{ij'}, y^c_{ij'1:T_{j'}}, f_{ij'1:L}, y^p_{ij'}\}$, the collection of past journey data by $\mathcal{H}_i = \bigcup\limits_{j' \in \mathcal{J}(i)} \mathcal{H}_{i,j'}$, the set of global parameters by Φ,[6] and all training data by $\mathcal{D}$.

We update the posterior of preferences for focal journey j, $\boldsymbol{\beta}_{ij}$, by

$$\begin{aligned} p(\boldsymbol{\beta}_{ij} | \mathbf{q}_{ij}, y^c_{ij1:t}, \mathcal{L}_{ijt}, \mathcal{H}_i, \mathcal{D}) &= \int p(\boldsymbol{\beta}_{ij} | \mathbf{q}_{ij}, y^c_{ij1:t}, \mathcal{L}_{ijt}, \mathcal{H}_i, \Phi) \cdot p(\Phi | \mathcal{D}, \mathbf{q}_{ij}, y^c_{ij1:t}, \mathcal{L}_{ijt}) d\Phi \\ &\approx \int p(\boldsymbol{\beta}_{ij} | \mathbf{q}_{ij}, y^c_{ij1:t}, \mathcal{L}_{ijt}, \mathcal{H}_i, \Phi) \cdot p(\Phi | \mathcal{D}) d\Phi, \end{aligned} \tag{8}$$

where $p(\Phi|\mathcal{D})$ is the posterior distribution of the global parameters given the training data. We expand the left term in (8), by drawing customer stable preferences, context-specific parameters, focal context membership, and past journeys contexts and marginalizing them,

$$\begin{aligned} p(\boldsymbol{\beta}_{ij} | \mathbf{q}_{ij}, &y^c_{ij1:t}, \mathcal{L}_{ijt}, \mathcal{H}_i, \Phi) = \\ &\int p(\boldsymbol{\beta}_{ij}, \boldsymbol{\mu}_i, \boldsymbol{\gamma}_j, \boldsymbol{\gamma}_{-j}, z_{ij}, \mathbf{z}_{i,-j} | \mathbf{q}_{ij}, y^c_{ij1:t}, \mathcal{L}_{ijt}, \mathcal{H}_i, \Phi) \cdot d\boldsymbol{\mu}_i \cdot d\boldsymbol{\gamma}_j \cdot d\boldsymbol{\gamma}_{-j} \cdot dz_{ij} \cdot d\mathbf{z}_{i,-j}, \end{aligned} \tag{9}$$

[6]Note that the global parameters are $\Phi = \left[\Sigma, \eta, a, d, \{\pi_c, \theta_c\}_{c=1}^C, \{\alpha^0_\ell, \boldsymbol{\alpha}^w_\ell, \boldsymbol{\alpha}^\beta_\ell\}_{\ell \in \{1,\ldots,L\}}\right]$.

where $\boldsymbol{\gamma}_{-j} = \{\boldsymbol{\gamma}_{j'}\}_{j' \in \mathcal{J}(i)}$. Finally, noting that $\boldsymbol{\gamma}_j = \begin{pmatrix} \boldsymbol{\omega}_j \\ \boldsymbol{\rho}_j \end{pmatrix}$ we can write this posterior as being proportional to the joint density

$$\begin{aligned}
&p(\boldsymbol{\beta}_{ij}, \boldsymbol{\mu}_i, \boldsymbol{\gamma}_j, \boldsymbol{\gamma}_{-j}, z_{ij}, \mathbf{z}_{i,-j} | \mathbf{q}_{ij}, y^c_{ij1:t}, \mathcal{L}_{ijt}, \mathcal{H}_i, \Phi) \\
&\quad \propto \quad p(\boldsymbol{\beta}_{ij}, \boldsymbol{\mu}_i, \boldsymbol{\gamma}_j, \boldsymbol{\gamma}_{-j}, z_{ij}, \mathbf{z}_{i,-j}, \mathbf{q}_{ij}, y^c_{ij1:t}, \mathcal{L}_{ijt}, \mathcal{H}_i |, \Phi) \\
&\quad = \quad p(\mathbf{q}_{ij} | \boldsymbol{\omega}_j) \cdot p(y^c_{ij1:t} | \boldsymbol{\mu}_i, \boldsymbol{\rho}_j, \eta) \cdot p(\mathcal{L}_{ijt} | \boldsymbol{\mu}_i, \boldsymbol{\rho}_j, \boldsymbol{\alpha}^0, \boldsymbol{\alpha}^w, \boldsymbol{\alpha}^\beta) \cdot \mathbf{1}\{\boldsymbol{\beta}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{\rho}_j\} \\
&\qquad \cdot \prod_{j' \in \mathcal{J}(i)} p(\mathcal{H}_{ij'} | \boldsymbol{\mu}_i, \boldsymbol{\gamma}_{-j}, \Phi) \\
&\qquad \cdot p(\boldsymbol{\mu}_i | \Sigma) \cdot p\left(z_{ij}, \boldsymbol{\gamma}_j | a, d, \{\pi_c, \theta_c\}_{c=1}^{\widetilde{C}}\right) \cdot \prod_{j' \in \mathcal{J}(i)} p\left(z_{ij'}, \boldsymbol{\gamma}_{j'} | a, d, \{\pi_c, \theta_c\}_{c=1}^{\widetilde{C}}\right),
\end{aligned} \tag{10}$$

where $\widetilde{C}$ is the number of contexts (which is a latent variable, and thus, it is drawn from the posterior $p(\Phi|\mathcal{D})$).

We update those parameters using steps 1, 2, 3, and 4 exactly as shown in Web Appendix C, and we adapt steps 5, 6, and 7 (and add step 8) to allow for previously unobserved contexts to be drawn by:

5*. Draw context membership $z_j \in \{1 \ldots \widetilde{C}\} \cup \{\widetilde{C} + 1\}$ as follows

$$p(z_j = c | \cdot) \propto \quad \begin{cases} (n_c - d) \cdot \mathcal{P}_{jc}, & \text{if } c \leqslant \widetilde{C} \\ (a + d \cdot \widetilde{C}) \cdot \mathcal{P}^*_j & \text{if } c = \widetilde{C} + 1 \end{cases}$$

where $\mathcal{P}_{jc}$ as defined in step 5 in Web Appendix C, and

$$\mathcal{P}^*_j = \left(\prod_{m=1}^{M} p(q_{ijm} | \phi_{0m})\right) \cdot p\left(\tilde{u}_{ij} - \widetilde{\mathbf{X}}_{ij} \boldsymbol{\mu}_i | \widetilde{\mathbf{X}}_{ij}, \mu_0, V_0\right)$$

the product of posterior predictive likelihoods[7] such that

$$\begin{aligned}
p(q_{ijm} | \phi_{0m}) &= \int p(q_{ijm} | \theta^w_m) \cdot p(\theta^w_m | \phi_{0m}) d\theta^w_m \\
p\left(\tilde{u}_{ij} - \widetilde{\mathbf{X}}_{ij} \boldsymbol{\mu}_i | \widetilde{\mathbf{X}}_{ij}, \mu_0, V_0\right) &= \int p\left(\tilde{u}_{ij} - \widetilde{\mathbf{X}}_{ij} \boldsymbol{\mu}_i | \widetilde{\mathbf{X}}_{ij} \theta^\rho_j\right) \mathcal{N}(\theta^\rho | \mu, V_0) d\theta^\rho.
\end{aligned}$$

6*. If $z_{ij} = \widetilde{C} + 1$, then draw the query components of context location parameters $\theta^\omega_{\widetilde{C}+1}$ following step 6 in Web Appendix C.

7*. If $z_{ij} = \widetilde{C} + 1$, then draw the click-purchase context location parameters $\theta^\rho_{\widetilde{C}+1}$ following step 7 in Web Appendix C.

[7] As all prior-likelihood pairs are conditionally conjugate, these posterior predictive likelihoods have closed form.

8*: If $z_{ij} = \widetilde{C} + 1$, then update $\widetilde{C} = \widetilde{C} + 1$, and repeat the same steps for all $j' \in \mathcal{J}(c)$.

E Details on predicting future activity for unfinished journeys

E.1 Approximation of consideration probabilities (XGBoost)

To compute the second term in (15), consider the probability that a product k belongs to the consideration set at the end of the journey (i.e., right before purchase) given the clicks observed up to step t, and given preferences $\boldsymbol{\beta}_{ij}$ inferred before then, i.e., $p(k \in \mathcal{C}_{ij}|y^c_{ij1:t}, \boldsymbol{\beta}_{ij})$. If the product has been clicked before step t, the probability of it being considered is one. If the product has not been clicked before t, the probability of it being considered afterwards involves an infinite sum because we do not know how many more steps the journey will have. Theoretically, one could build a forward-looking search model to estimate the consideration set probabilities. Given how non-linear our journeys are and the large number of possible items to query, filter, and click on, such a model would be intractable. Additionally, in order to use our model to simulate future clicks and choices we would need to observe hypothetical choice sets that depend on the path being simulated. These choice sets are observed during training, but unobserved for journeys that are incomplete. Since customers can visit the website across multiple sessions within the same purchase journey, simulating choices also involves defining a process for drawing choice sets. For example, consider making after-query predictions on a journey that lasted a single step. When simulating choices, there is no data about what products may be available after the first step, especially if the simulated journey takes multiple steps.

To make such a forecast tractable and scalable, we approximate this probability by imputing the predictions from a flexible (reduced-form) model that, leveraging finished journeys in the training data, estimates the probability that a product will be added to the consideration set.

Specifically, we infer consideration given product characteristics $\mathbf{x}_{ijk}$ and preferences $\boldsymbol{\beta}_{ij}$ through a reduced-form predictor function $\hat{g}_{\mathcal{C}}(\mathbf{x}, \boldsymbol{\beta})$ such that,

$$p(k \in \mathcal{C}_{ij}|y^c_{ij1:t}, \boldsymbol{\beta}_{ij}) \approx \begin{cases} 1 & \text{if clicked on before, i.e., } \exists t' \leqslant t,\ y^c_{ijt'} = k, \\ \hat{g}_{\mathcal{C}}(\mathbf{x}_{ijk}, \beta_{ij}) & \sim . \end{cases} \tag{11}$$

Such a prediction function can be estimated using standard machine learning (ML) models trained using all displayed products in finished journeys of the training data, for which we precisely observe whether each product was added to the consideration set. In our application, we use a binary XGBoost model to estimate the function $\hat{g}_{\mathcal{C}}(\mathbf{x}, \boldsymbol{\beta})$. As features for the XGBoost model, we use the exact observed product attributes used in the purchase model ($\mathbf{x}_{ijk}$) as well as draws from the posterior distribution of (individual-level) customer preferences ($\boldsymbol{\beta}_{ij}$), obtained from the main model when estimated using the same journeys in the training data. Note that adding customer preferences as features enriches the ML model predictions as those capture the unobserved individual-level preferences.

In our empirical application, consideration is operationalized slightly differently for the two types of flights: one-way and roundtrip. For one-way itineraries, the details page is shown after a single click on a one-way results page, the moment at which we assume the flight is being considered. For roundtrip itineraries, on the other hand, the customer must click on the outbound component of the flight (on an outbound results page), *and* on the inbound (return) component of the flight in order to see the details page and for the product to be considered. Accordingly, we train three different models, each aiming at a different prediction task: One that predicts consideration for one-way flights ($\hat{g}_{ow}$), another one that predicts whether the outbound component of a roundtrip flight is considered ($\hat{g}_{out}$), and another one that predicts, conditional on the outbound component being considered, whether the inbound component is also considered ($\hat{g}_{in}$).

Following (11), we compute the consideration probabilities given whether the customer has clicked on the itinerary, or a portion of the itinerary. That is,

$$p(k \in \mathcal{C}_{ij}|\text{One-way}, y^c_{ij1:t}, \boldsymbol{\beta}_{ij}) \approx \begin{cases} 1 & \text{if flight was clicked on before} \\ \hat{g}_{ow}(\mathbf{x}_{ijk}, \beta_{ij}) & \text{if flight has not been clicked on yet,} \end{cases}$$

$$p(k \in \mathcal{C}_{ij}|\text{Roundtrip}, y^c_{ij1:t}, \boldsymbol{\beta}_{ij}) \approx \begin{cases} 1 & \text{if both legs were clicked on} \\ 1 \cdot \hat{g}_{in}(\mathbf{x}_{ijk}, \beta_{ij}) & \text{if only outbound leg was clicked on} \\ \hat{g}_{out}(\mathbf{x}_{ijk}, \beta_{ij}) \cdot \hat{g}_{in}(\mathbf{x}_{ijk}, \beta_{ij}) & \text{if no leg has been clicked on.} \end{cases}$$

We estimate tree-based classifiers (XGBoost and Random Forest) to predict consideration in hold out journeys. We train such models using the data from the training sample (including clicks as the dependent variable and the product attributes as features) as well as draws from the posterior distribution of the vector of preferences (which are included as additional features in our classifier).

Because the parameters $\boldsymbol{\beta}_{ij}$ are estimated in a Bayesian manner (i.e., we don't have a point estimate but a posterior distribution), we draw a sample of 50 draws from the posterior distribution of $\boldsymbol{\beta}_{ij}$ when training the consideration of each journey. Specifically, for each product k in a journey, we create 50 observations, each with a feature vector concatenating the vector of product attributes, $\mathbf{x}_{ijk}$, and the drawn preferences $\widetilde{\boldsymbol{\beta}}_{ijd}$. We sample 1,000,000 observations ($\sim$1% of total) to train the classifiers. We use one-way observations to train $\hat{g}_{ow}$; and roundtrip observations to train $\hat{g}_{out}$. To train $\hat{g}_{in}$, we only use roundtrip observations such that the outbound leg of the corresponding itinerary was clicked on.[8] We use as binary outcomes whether the corre-

[8] Arguably, there could be selection bias affecting our sample as we would make predictions for those products not clicked on yet based only on those clicked on the outbound leg. However, we argue that this approach is the most sensible given the task at hand. First, any potential selection bias should hurt the out-of-sample performance, and, thus, be captured by the out-of-sample performance of the predictions of the whole model. Second, those predictions should only be relevant for products that had their outbound leg clicked on, or that the outbound model predicts will be clicked on. Therefore, even if predictions are off for products that are unlikely to be clicked on, they are already captured by $\widehat{g}_{out}$.

sponding product of each observation was clicked on during the journey,[9] and the corresponding cross-entropy loss (i.e., binary logistic) to train the models.

Table E.2: Performance of XGBoost consideration predictors.

	Consideration				
Model	Balanced accuracy	Precision	Recall	F1	AUC
One-way ($\hat{g}_{ow}$)					
XGBoost	0.2287	0.3893	0.0680	0.1158	0.9064
Random Forest	0.3403	0.6486	0.0320	0.0610	0.6898
Outbound ($\hat{g}_{out}$)					
XGBoost	0.9598	0.9406	0.9789	0.9594	0.9958
Random Forest	0.8027	0.8304	0.7749	0.8017	0.9593
Inbound ($\hat{g}_{in}$)					
XGBoost	0.3488	0.5482	0.1494	0.2348	0.9233
Random Forest	0.3928	0.6879	0.0977	0.1711	0.7737

We use a 80%-20% training/test split, and ten-fold cross-validation on the training sample over a grid to tune the hyperparameters of each classifier (e.g., the learning rate and the maximum depth of the trees for the XGBoost). Table E.2 shows the performance of both the XGBoost and the Random Forest on each prediction task. Because the XGBoost overall accuracy metrics (F1 and AUC) are superior in all tasks, we use the results of the XGBoost when augmenting consideration sets.

[9] For the outbound leg model, we use as an outcome whether the product has the same exact outbound leg as any product that was clicked on during the journey. That is, if an outbound leg is clicked on within a results page, all returning flights displayed on the next page (which share the same already-clicked outbound leg) are defined as positive labels for the predictive model.

E.2 Computing purchase probabilities

Following (11), we can now approximate the conditional purchase probabilities in (15) using a Monte Carlo approximation where we draw consideration sets. In each iteration of our MC simulation, we form each consideration set by first including all products that have been clicked on up to that point. Subsequently, for all remaining products, we add them to the consideration set with a probability given by $\hat{g}_{\mathcal{C}}(\mathbf{x}, \boldsymbol{\beta})$. Finally, once we have drawn the consideration set in each iteration of our simulation, we compute the purchase probabilities given each consideration set.[10] All other products that do not belong to the consideration have a null purchase probability. We outline such procedure in Algorithm 1 where we compute the purchase probabilities given a draw from the posterior distribution $p(\boldsymbol{\beta}_{ij}|\mathbf{q}_{ij}, y^c_{ij1:t}, \mathcal{L}_{ijt})$.

Algorithm 1 Computing purchase probabilities

Input
- A vector of preferences $\boldsymbol{\beta}_{ij}$
- A set of products with at least one click $\mathcal{C}^{obs}_{ij} = \{k| \ \exists t' \leqslant t, \ y^c_{ijt'} = k\}$
- Number of samples S for the Monte Carlo approximation
- Trained predictor function $\hat{g}_{\mathcal{C}}(\mathbf{x}, \boldsymbol{\beta})$

Output
- $p(y^p_{ij}|y^c_{ij1:t}, \boldsymbol{\beta}_{ij})$

Procedure

for all $s \leftarrow 1 : S$ **do**
    Initialize consideration set $\mathcal{C}_{ij} \leftarrow \mathcal{C}^{obs}_{ij}$
    for all $k \notin \mathcal{C}^{obs}_{ij}$ **do**
        Draw $u \sim U(0, 1)$
        if $u \leqslant \hat{g}_{\mathcal{C}}(\mathbf{x}_{ijk}, \boldsymbol{\beta}_{ij})$ **then**
            $\mathcal{C}_{ij} \leftarrow \mathcal{C}_{ij} \cup \{k\}$
        end if
    end for
    Compute $p_s = p(y^p_{ij}|\mathcal{C}_{ij}, \boldsymbol{\beta}_{ij})$ using GLK simulator and Equation (6)
end for
Return $p(y^p_{ij}|y^c_{ij1:t}, \boldsymbol{\beta}_{ij}) \approx \frac{1}{S}\sum_{s=1}^{S} p_s$

---

[10] Specifically, we use the GHK-algorithmGeweke (1991) to approximate purchase probabilities from a multinomial probit model given a consideration set.

F Empirical application: Additional summary statistics

F.1 Product attributes

Table F.3: Summary statistics of product attributes in page results

Product attribute	Mean	SD	Quantiles 5%	50%	95%
Product level attributes					
Price	1,547	3,269	196	751	5,320
Cheapest price per journey	698	1,526	98	401	2,117
Outbound level attributes					
Length of trip (hours)	11.28	8.49	2.05	8.42	28.60
Shortest length of trip per journey (hours)	5.86	5.05	1.25	4.07	17.08
Number of stops: Non stop	0.20	.	0	0	1
Number of stops: One stop	0.59	.	0	1	1
Number of stops: 2+ stops	0.21	.	0	0	1
Alliance: Alaska Airlines	0.04	.	0	0	0
Alliance: Frontier	0.01	.	0	0	0
Alliance: JetBlue	0.03	.	0	0	0
Alliance: Multiple alliances	0.07	.	0	0	1
Alliance: Other – No alliance	0.07	.	0	0	1
Alliance: OneWorld (American)	0.27	.	0	0	1
Alliance: Skyteam (Delta)	0.27	.	0	0	1
Alliance: Spirit	0.02	.	0	0	0
Alliance: Star Alliance (United)	0.23	.	0	0	1
Dep. time: Early morning (0:00am - 4:59am)	0.04	.	0	0	0
Dep. time: Morning (5:00am – 11:59am)	0.47	.	0	0	1
Dep. time: Afternoon (12:00pm - 5:59pm)	0.31	.	0	0	1
Dep. time: Evening (6:00pm - 11:59pm)	0.18	.	0	0	1
Arr. time: Early morning (0:00am - 4:59am)	0.05	.	0	0	0
Arr. time: Morning (5:00am – 11:59am)	0.24	.	0	0	1
Arr. time: Afternoon (12:00pm - 5:59pm)	0.34	.	0	0	1
Arr. time: Evening (6:00pm - 11:59pm)	0.37	.	0	0	1
Inbound level attributes					
Length of trip (hours)	11.08	9.02	1.83	7.92	29.50
Shortest length of trip per journey (hours)	6.17	5.31	1.25	4.27	17.75
Number of stops: Non stop	0.19	.	0	0	1
Number of stops: One stop	0.70	.	0	1	1
Number of stops: 2+ stops	0.11	.	0	0	1
Alliance: Alaska Airlines	0.02	.	0	0	0
Alliance: Frontier	0.02	.	0	0	0
Alliance: JetBlue	0.02	.	0	0	0
Alliance: Multiple alliances	0.02	.	0	0	0
Alliance: Other – No alliance	0.07	.	0	0	1
Alliance: OneWorld (American)	0.51	.	0	1	1
Alliance: Skyteam (Delta)	0.13	.	0	0	1
Alliance: Spirit	0.05	.	0	0	1
Alliance: Star Alliance (United)	0.15	.	0	0	1
Dep. time: Early morning (0:00am - 4:59am)	0.03	.	0	0	0
Dep. time: Morning (5:00am – 11:59am)	0.65	.	0	1	1
Dep. time: Afternoon (12:00pm - 5:59pm)	0.18	.	0	0	1
Dep. time: Evening (6:00pm - 11:59pm)	0.14	.	0	0	1
Arr. time: Early morning (0:00am - 4:59am)	0.04	.	0	0	0
Arr. time: Morning (5:00am – 11:59am)	0.55	.	0	1	1
Arr. time: Afternoon (12:00pm - 5:59pm)	0.19	.	0	0	1
Arr. time: Evening (6:00pm - 11:59pm)	0.23	.	0	0	1

F.2 Filter construction and summary statistics

As mentioned in the main manuscript, the focal company did not collect the action of "filtering" directly. Rather, we infer such a behavior from the flight results we observe in the data. Specifically, we construct filter data conservatively in the following manner: (1) We infer that a filter was applied if all product results on a page have the same level on a product attribute (e.g., non-stop) and this does not occur in the first page of results.[11] (2) We allow multiple filters on a page as long as they belong to different attributes (e.g., American Airlines and non-stop).

Similar to the click and purchase data, airline data in filters is equally sparse, so we aggregate them into filters at the alliance level. That said, we still infer whether a filter was applied on a page using the airline data, as customers could only apply filters at the airline level and not at the alliance level. For example, if a page contains results from multiple OneWorld airlines (e.g., American Airlines and British Airways results), we do not define those results as resulting from a filter, as the platform did not allow customers to filter specifically on alliances. However, we define a filter on the OneWorld alliance if all flights belong to a single airline that belongs to the OneWorld alliance (e.g., all flights American Airlines or all flights British Airways).

Table F.4 shows, per attribute and level, the percentage of first-party journeys where a filter was applied.

Table F.4: Percentage of journeys with filters in attributes.

Attribute	**Level**	**Proportion journeys filtered**	
		Mean	s.e.
Alliance	OneWorld	0.020	0.001
	Skyteam	0.016	0.001
	Star Alliance	0.017	0.001
	Alaska Airlines	0.003	0.000
	Frontier	0.001	0.000
	JetBlue	0.006	0.000
	Spirit	0.001	0.000
	OTHER_NO_ALLIANCE	0.008	0.001
Stops	Non-stop	0.138	0.002
	One stop	0.038	0.001
Departure time	Early morning (0:00am - 4:59am)	0.004	0.000
	Morning (5:00am - 11:59am)	0.032	0.001
	Afternoon (12:00pm - 5:59pm)	0.027	0.001
	Evening (6:00pm - 11:59pm)	0.028	0.001
Arrival time	Early morning (0:00am - 4:59am)	0.002	0.000
	Morning (5:00am - 11:59am)	0.018	0.001
	Afternoon (12:00pm - 5:59pm)	0.019	0.001
	Evening (6:00pm - 11:59pm)	0.021	0.001

[11] Because the website does not filter by default, a constant attribute on the first page reflects limited supply, not a filtering constraint.

G Empirical application: Additional results

G.1 Context-specific parameter estimates

Table G.5: Posterior mean of location click and purchase parameters. Contexts 1-11

	Context										
Parameter	1	2	3	4	5	6	7	8	9	10	11
Query											
Is it roundtrip? Yes	0.24	0.91	0.90	0.99	0.97	0.93	0.95	0.99	0.20	0.96	0.97
Is it domestic? (within EU is domestic) Yes	1.00	1.00	1.00	0.00	0.02	1.00	0.00	1.00	0.08	0.00	0.03
Flying from international airport? Yes	0.55	0.74	0.48	0.87	0.74	0.50	0.94	0.62	0.90	0.90	0.90
Market: US Domestic	0.98	0.95	0.97	0.00	0.00	0.97	0.00	0.93	0.00	0.00	0.00
Market: US Overseas	0.00	0.00	0.00	0.83	0.01	0.00	0.72	0.00	0.00	0.78	0.01
Market: Non-US across continent	0.00	0.00	0.00	0.15	0.05	0.00	0.26	0.00	0.09	0.20	0.08
Market: Non-US within continent	0.00	0.02	0.00	0.01	0.07	0.00	0.01	0.04	0.19	0.01	0.10
Market: US North America	0.02	0.02	0.03	0.00	0.88	0.02	0.00	0.03	0.72	0.00	0.81
Type of location searched: Airport	0.92	0.89	0.93	0.88	0.90	0.88	0.89	0.91	0.90	0.85	0.81
Type of location searched: Both	0.02	0.01	0.02	0.02	0.06	0.04	0.04	0.02	0.07	0.04	0.07
Type of location searched: City	0.05	0.10	0.05	0.10	0.04	0.08	0.07	0.07	0.03	0.12	0.12
Trip distance (1000s kms)	1.87	2.53	1.84	9.72	2.71	0.80	9.71	2.36	2.53	8.96	2.77
More than one adult? Yes	0.17	0.39	0.24	0.28	0.46	0.22	0.18	0.50	0.26	0.26	0.47
Traveling with kids? Yes	0.04	0.13	0.06	0.03	0.17	0.05	0.11	0.17	0.10	0.07	0.12
Is it summer season? Yes	0.43	0.39	0.42	0.00	0.26	0.36	0.21	0.00	0.62	0.98	0.21
Holiday season? Yes	0.00	0.00	0.01	0.06	0.07	0.01	0.00	0.32	0.00	0.00	0.09
Does stay include a weekend? Yes	0.15	0.89	0.99	1.00	0.97	0.83	1.00	0.99	0.25	1.00	0.90
Length of stay (only RT) (days)	2.30	5.25	5.63	14.77	9.83	3.88	45.67	6.16	2.81	13.16	7.40
Searching on weekend? Yes	0.19	0.21	0.21	0.25	0.24	0.19	0.26	0.17	0.24	0.27	0.20
Searching during work hours? Yes	0.52	0.52	0.52	0.51	0.55	0.54	0.37	0.59	0.42	0.43	0.60
Time in advance to buy (days)	23.95	58.21	41.18	107.77	81.79	37.34	53.67	111.04	29.07	39.38	92.67
Preferences											
Intercept Search: OW Search	-0.02	-0.28	0.01	-0.14	0.04	-0.19	0.06	0.09	-0.06	-0.18	-0.16
Intercept Search: RT Outbound	-0.22	-0.51	-0.13	-0.70	-0.13	-0.49	-0.10	-0.09	-0.24	-0.41	-0.45
Intercept Search: RT Inbound	-0.03	-0.11	-0.26	-0.18	-0.01	-0.13	-0.11	-0.15	-0.06	-0.10	-0.04
Intercept Click: OW Search	-0.81	-0.37	-0.61	0.04	0.08	-0.09	-0.08	-0.09	-0.36	-0.02	-0.03
Intercept Click: RT Outbound	-0.28	-0.17	-0.18	-0.14	-0.18	-0.43	-0.31	-0.61	-0.04	-0.21	-0.43
Intercept Click: RT Inbound	0.05	-0.18	0.19	0.28	0.35	-0.15	0.43	0.08	-0.02	0.15	-0.13
Price	-0.19	0.02	-0.37	-0.19	-0.18	0.15	-0.34	-0.15	-0.03	-0.06	0.32
Length of trip (hours)	-0.59	-0.54	-0.90	-0.59	-0.85	-0.20	-0.69	-0.46	-0.49	-0.21	-0.09
Number of stops: Non stop	0.11	0.43	0.60	0.31	0.60	0.14	0.35	0.18	-0.02	0.22	0.03
Number of stops: 2+ stops	-0.33	-0.15	-0.28	-0.46	-0.12	-0.06	-0.40	-0.10	-0.10	-0.25	-0.05
Alliance: Skyteam (Delta)	-0.02	-0.12	-0.18	-0.09	-0.14	-0.09	-0.01	-0.06	-0.12	0.03	-0.10
Alliance: Star Alliance (United)	-0.10	-0.19	-0.21	0.14	0.12	-0.16	-0.06	-0.09	0.02	-0.06	-0.06
Alliance: Alaska Airlines	-0.06	-0.08	0.01	0.02	-0.10	-0.01	0.13	-0.01	-0.07	-0.01	-0.04
Alliance: Spirit	-0.21	-0.02	-0.21	0.01	0.00	0.04	-0.06	-0.05	-0.01	0.00	0.01
Alliance: JetBlue	-0.02	0.23	-0.01	0.00	0.08	0.13	-0.13	-0.03	0.04	0.04	0.02
Alliance: Frontier	-0.11	0.06	-0.03	-0.03	0.00	0.02	-0.02	0.01	0.06	0.01	0.00
Alliance: Other – No alliance	-0.09	-0.05	-0.04	0.05	0.02	-0.03	0.12	0.12	-0.14	-0.06	-0.04
Alliance: Multiple alliances	-0.13	-0.07	-0.10	0.02	-0.08	-0.02	-0.09	-0.04	-0.09	0.00	0.03
Outbound dep. time: Early morning (0:00am - 4:59am)	-0.16	-0.01	0.07	0.05	0.11	0.01	-0.08	0.01	-0.03	-0.01	-0.02
Outbound dep. time: Afternoon (12:00pm - 5:59pm)	-0.08	-0.26	-0.09	0.03	0.01	-0.03	-0.14	-0.06	-0.04	-0.10	-0.07
Outbound dep. time: Evening (6:00pm - 11:59pm)	-0.27	-0.21	-0.07	-0.07	-0.10	-0.04	-0.11	-0.01	-0.03	-0.07	-0.07
Outbound arr. time: Early morning (0:00am - 4:59am)	-0.11	-0.14	-0.20	0.00	-0.05	-0.04	-0.11	-0.05	-0.03	-0.05	-0.02
Outbound arr. time: Afternoon (12:00pm - 5:59pm)	0.10	0.13	0.19	-0.11	0.04	-0.11	-0.05	-0.02	-0.09	-0.12	-0.08
Outbound arr. time: Evening (6:00pm - 11:59pm)	-0.01	-0.17	0.12	-0.08	-0.12	0.01	-0.14	-0.14	-0.07	-0.01	-0.06
Inbound dep. time: Early morning (0:00am - 4:59am)	-0.08	0.11	0.00	0.10	-0.01	0.12	-0.07	-0.15	-0.02	0.11	0.10
Inbound dep. time: Afternoon (12:00pm - 5:59pm)	0.10	0.23	0.35	0.05	0.15	-0.01	0.07	0.09	-0.01	0.00	-0.03
Inbound dep. time: Evening (6:00pm - 11:59pm)	0.06	-0.03	0.06	0.05	-0.04	-0.01	0.12	-0.01	0.01	0.03	-0.01
Inbound arr. time: Early morning (0:00am - 4:59am)	-0.11	0.05	0.00	0.02	0.05	0.07	-0.13	-0.07	0.00	0.11	0.08
Inbound arr. time: Afternoon (12:00pm - 5:59pm)	0.01	0.03	0.14	-0.04	0.09	0.00	0.09	0.03	-0.05	-0.02	-0.06
Inbound arr. time: Evening (6:00pm - 11:59pm)	0.20	0.21	0.44	-0.01	0.17	-0.01	-0.05	0.09	0.00	0.02	-0.04

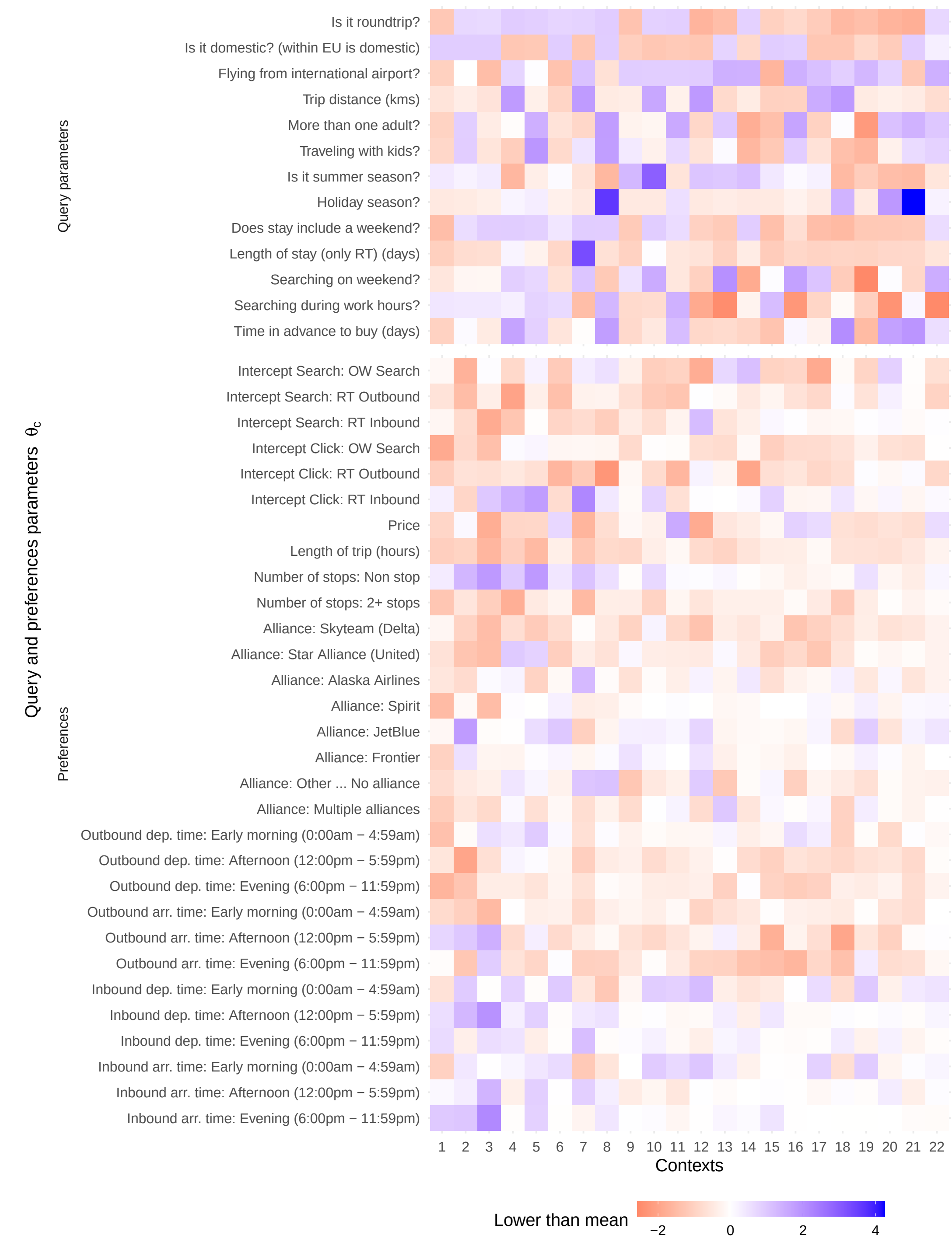
Table G.6: Posterior mean of location click and purchase parameters. Contexts 12-22

	Context										
Parameter	12	13	14	15	16	17	18	19	20	21	22
Query											
Is it roundtrip? Yes	0.09	0.16	0.96	0.31	0.37	0.27	0.13	0.17	0.09	0.05	0.92
Is it domestic? (within EU is domestic) Yes	0.00	0.94	0.19	1.00	0.97	0.00	0.00	0.18	0.06	1.00	0.71
Flying from international airport? Yes	0.90	1.00	0.99	0.44	1.00	0.94	0.89	0.97	0.88	0.52	1.00
Market: US Domestic	0.00	0.05	0.00	0.98	0.15	0.00	0.00	0.00	0.00	0.96	0.01
Market: US Overseas	0.87	0.00	0.00	0.00	0.00	0.66	0.79	0.01	0.00	0.00	0.00
Market: Non-US across continent	0.11	0.00	0.06	0.00	0.00	0.30	0.19	0.15	0.06	0.00	0.01
Market: Non-US within continent	0.01	0.93	0.29	0.01	0.74	0.04	0.01	0.43	0.36	0.02	0.97
Market: US North America	0.00	0.02	0.65	0.01	0.11	0.00	0.00	0.41	0.58	0.02	0.01
Type of location searched: Airport	0.84	0.76	0.83	0.90	0.76	0.85	0.88	0.87	0.79	0.91	0.84
Type of location searched: Both	0.01	0.10	0.08	0.04	0.03	0.04	0.03	0.09	0.16	0.06	0.13
Type of location searched: City	0.15	0.14	0.09	0.05	0.22	0.11	0.10	0.04	0.06	0.03	0.03
Trip distance (1000s kms)	9.81	1.16	2.41	0.50	0.59	8.70	9.84	2.37	2.72	2.39	1.37
More than one adult? Yes	0.19	0.40	0.08	0.13	0.48	0.17	0.29	0.03	0.42	0.45	0.41
Traveling with kids? Yes	0.06	0.09	0.01	0.03	0.13	0.05	0.02	0.01	0.07	0.11	0.12
Is it summer season? Yes	0.57	0.56	0.59	0.43	0.37	0.40	0.01	0.10	0.03	0.02	0.22
Holiday season? Yes	0.00	0.01	0.00	0.00	0.02	0.00	0.15	0.00	0.19	0.37	0.06
Does stay include a weekend? Yes	0.30	0.25	0.99	0.16	0.41	0.15	0.11	0.23	0.24	0.25	0.90
Length of stay (only RT) (days)	6.53	2.92	8.55	1.56	3.92	3.09	3.40	3.17	4.03	4.21	6.96
Searching on weekend? Yes	0.18	0.29	0.14	0.22	0.28	0.26	0.17	0.12	0.22	0.18	0.27
Searching during work hours? Yes	0.34	0.29	0.47	0.58	0.31	0.41	0.48	0.41	0.30	0.50	0.29
Time in advance to buy (days)	28.34	30.26	25.97	14.45	60.45	46.61	121.05	8.12	109.18	116.00	73.98
Preferences											
Intercept Search: OW Search	-0.30	0.12	0.19	-0.16	-0.15	-0.31	-0.02	-0.15	0.14	-0.01	-0.12
Intercept Search: RT Outbound	0.01	-0.04	-0.17	-0.08	-0.23	-0.30	0.03	-0.22	0.09	-0.02	-0.34
Intercept Search: RT Inbound	0.17	-0.08	-0.05	0.02	0.01	-0.03	-0.02	0.01	0.02	-0.01	0.01
Intercept Click: OW Search	-0.31	-0.34	-0.06	-0.46	-0.35	-0.34	-0.28	-0.14	-0.29	-0.32	-0.01
Intercept Click: RT Outbound	0.06	-0.06	-0.52	-0.19	-0.16	-0.24	-0.20	0.02	-0.04	0.02	-0.23
Intercept Click: RT Inbound	0.01	0.00	0.02	0.16	-0.04	-0.04	0.09	-0.03	0.03	-0.04	0.02
Price	-0.38	-0.11	-0.08	-0.04	0.17	0.13	-0.14	-0.15	-0.13	-0.15	0.13
Length of trip (hours)	-0.45	-0.53	-0.33	-0.23	-0.22	-0.08	-0.35	-0.35	-0.38	-0.29	-0.15
Number of stops: Non stop	0.02	0.05	-0.01	-0.05	-0.11	-0.06	-0.04	0.18	-0.06	-0.13	0.06
Number of stops: 2+ stops	-0.15	-0.09	-0.09	-0.09	-0.03	-0.12	-0.30	-0.11	-0.01	-0.07	-0.03
Alliance: Skyteam (Delta)	-0.16	-0.05	-0.07	-0.04	-0.16	-0.12	-0.09	-0.05	-0.08	-0.07	-0.03
Alliance: Star Alliance (United)	-0.07	0.02	-0.07	-0.16	-0.13	-0.18	-0.09	-0.01	-0.03	-0.01	-0.04
Alliance: Alaska Airlines	0.02	-0.03	0.04	-0.07	-0.03	-0.02	0.03	-0.05	0.02	-0.06	-0.03
Alliance: Spirit	0.00	-0.03	-0.02	0.00	0.00	0.02	-0.02	0.04	-0.04	0.02	0.02
Alliance: JetBlue	0.10	-0.03	-0.02	-0.01	-0.02	0.02	-0.10	0.12	-0.08	0.03	0.06
Alliance: Frontier	0.06	-0.04	-0.01	-0.02	-0.04	0.00	-0.01	0.03	0.01	-0.03	0.00
Alliance: Other – No alliance	0.10	-0.13	-0.01	0.02	-0.11	-0.03	-0.05	-0.08	-0.01	-0.03	-0.03
Alliance: Multiple alliances	-0.09	0.12	-0.07	0.02	-0.01	0.02	-0.12	0.04	-0.01	-0.03	0.00
Outbound dep. time: Early morning (0:00am - 4:59am)	-0.02	0.02	-0.04	-0.02	0.07	0.04	-0.12	-0.01	-0.10	0.00	-0.02
Outbound dep. time: Afternoon (12:00pm - 5:59pm)	-0.04	0.00	-0.10	-0.14	-0.08	-0.10	-0.12	-0.09	-0.08	-0.11	-0.01
Outbound dep. time: Evening (6:00pm - 11:59pm)	-0.06	-0.17	0.00	-0.16	-0.19	-0.17	-0.06	-0.07	-0.04	-0.12	-0.04
Outbound arr. time: Early morning (0:00am - 4:59am)	-0.13	-0.09	-0.06	0.00	-0.04	-0.05	-0.06	-0.01	-0.08	-0.10	0.00
Outbound arr. time: Afternoon (12:00pm - 5:59pm)	-0.03	0.04	-0.05	-0.23	-0.04	-0.10	-0.26	-0.08	-0.14	-0.01	0.00
Outbound arr. time: Evening (6:00pm - 11:59pm)	-0.13	-0.14	-0.18	-0.19	-0.22	-0.12	-0.18	0.05	-0.10	-0.09	-0.02
Inbound dep. time: Early morning (0:00am - 4:59am)	0.15	-0.04	-0.07	-0.06	0.00	0.07	-0.09	0.12	-0.04	0.05	0.06
Inbound dep. time: Afternoon (12:00pm - 5:59pm)	-0.02	0.06	-0.06	0.08	-0.02	-0.02	0.01	0.00	0.02	-0.01	0.03
Inbound dep. time: Evening (6:00pm - 11:59pm)	-0.04	0.02	0.03	-0.01	-0.01	-0.01	0.04	-0.03	0.03	-0.02	-0.01
Inbound arr. time: Early morning (0:00am - 4:59am)	0.12	0.04	-0.03	0.00	0.00	0.09	-0.08	0.10	-0.03	0.01	0.02
Inbound arr. time: Afternoon (12:00pm - 5:59pm)	0.00	-0.01	0.00	0.00	0.00	-0.02	0.01	-0.01	0.04	-0.04	0.01
Inbound arr. time: Evening (6:00pm - 11:59pm)	0.00	0.04	0.02	0.10	0.00	0.00	0.00	0.00	0.00	-0.02	-0.02

G.2 Relative differences across contexts (all variables)

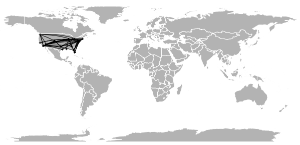
We normalize the location parameters to account for how they vary across contexts and how much uncertainty their posterior have. First, for each context c, we compute the posterior mean of each location parameter θ_c. Second, we compare these location parameters with the population mean level of those same parameters, but now we include query parameters as well. We subtract these two to measure whether contexts are *above or below average* on each of the query parameters and click and purchase preferences. Finally, we normalize these differences by dividing by the square root of the posterior variance across journeys. This variance is composed by two terms (similar to ANOVA): (1) the within-context posterior variance of each θ_c, which measures the posterior uncertainty of each location parameter θ_c; and (2) the across-context variance of all θ_c with respect to the population mean, which captures how much variance is explained by the differences between contexts. By normalizing the location parameters, we can now compare contexts with respect to whether they score higher or lower than average on each of the query parameters and preferences.

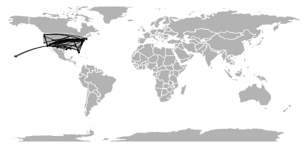
Figure G.2: Posterior mean of all context location parameters θ_c, relative to the average in the population. The top figure shows how each context deviates from the average with respect to the query variables. The bottom figure shows deviations with respect to the preference parameters. Blue (red) boxes mean positive (negative) deviation from the average in the population.

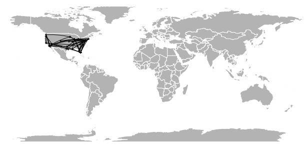


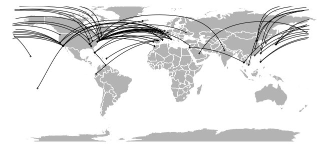
G.3 Top 50 routes per context: All contexts

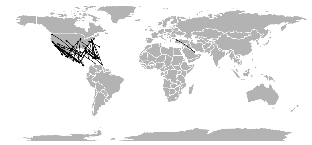
Figure G.3: Top 50 routes per context

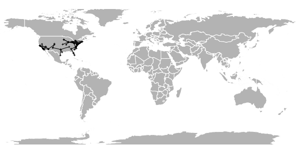

(a) Context 1


(b) Context 2

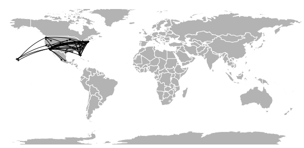

(c) Context 3

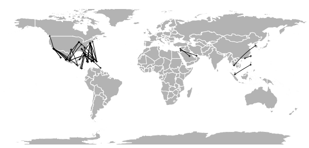

(d) Context 4


(e) Context 5

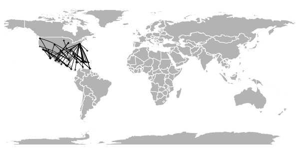

(f) Context 6


(g) Context 7

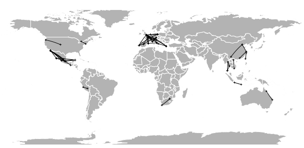

(h) Context 8

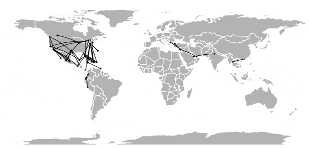

(i) Context 9

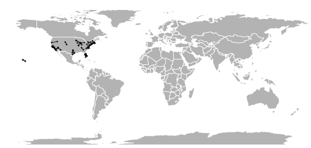

(j) Context 10

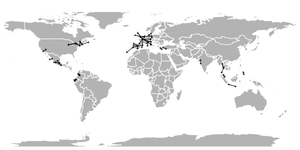

(k) Context 11


(l) Context 12

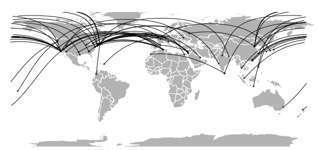

(m) Context 13

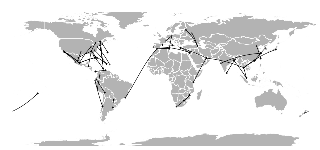

(n) Context 14

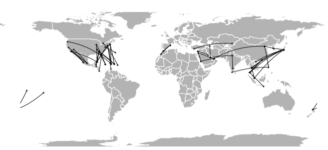

(o) Context 15

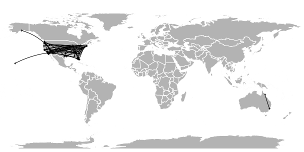

(p) Context 16

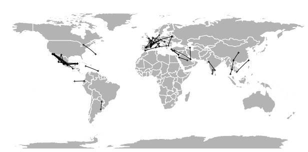

(q) Context 17


(r) Context 18


(s) Context 19


(t) Context 20


(u) Context 21


(v) Context 22

H Model predictive ability

To assess the predictive performance of the model, we focus on predicting whether a transaction occurs at the end of each (held-out) journey, and if so, which product is chosen. We compare our model with other established methodologies that have been extensively tested for such predictive tasks. We evaluate predictive validity of our model on two distinct occasions: immediately after the customer submits the query (i.e., before any clicks are observed) and once the customer has clicked on five occasions. These time points (i.e., after query and after clicks) dictate the available information for making predictions, both for our model and the benchmark models.[12]

H.1 Benchmarks of comparison

We compare our model to two commonly used machine learning classification methods, Random Forest (RF) and XGBoost. As these benchmarks can only be used to predict outcomes given a set of features, we create a comprehensive set of features that capture the information available (to the firm) at each prediction point. In addition to the attributes of the corresponding product, our set of features includes summary statistics from: (1) the query of the focal journey (capturing the attributes of the search), (2) the attributes of the products shown in the first page of the focal journey (capturing the attributes of the products presented to the consumer), (3) the clicks and filters (during the focal journey) up to the moment when the prediction is made (capturing what the customer has been clicking), and (4) the queries, product attributes, clicks, filters, and purchases from past journeys (capturing past behavior). (Please refer to Web Appendix H.2 for a detailed description of the features and the estimation of benchmarks.)

Machine learning classification models are often built for binary tasks. In our case, not only is the classification not binary, but the choice set size (the number of possible available flights to choose from) varies from one classification task (journey) to another. To address this complication, we create a series of binary classifications and use normalization to convert these to a multinomial choice task. Specifically, we train each benchmark model as a binary predictive model where each observation represents choosing one product in a particular journey (for each journey, we include an additional no-purchase “product” for the outside option). Before making predictions, we normalize the prediction scores per customer per journey such that the scores for all alternatives (in each journey) sum to one. To predict if the customer buys any product or does not buy, we label it as a purchase if the normalized score for the no-purchase alternative is lower than 0.5. For the product choice task, we normalize the predictive scores of each product by dividing by the sum of the scores of all products except the outside option (i.e., choice conditional on purchase) and label as the chosen product, the product with the highest score. For this evaluation, we only consider (held-out) journeys that ended up in a purchase, for which we observe the actual product choice.

[12] When a (held-out) journey has less than 5 clicks occasions, we use all clicks prior to purchase. Doing so not only provides a more conservative measure of the predictive improvements but also avoids selection biases due to shorter and longer journeys. Also, results are qualitatively similar when using different numbers of click occasions.

H.2 Estimation of Benchmark models

We describe in detail how the benchmark models are trained and how the binary prediction scores are normalized per journey. As both benchmark models are built for binary classification tasks (or multi-class classification tasks with a fixed set of classes across observations), we create a series of binary classifications and use normalization to convert these to a multinomial choice task (or varying choice sizes, depending on the consideration set of each journey).

Consider customer i, in journey j and the set $\mathcal{K}_j$ that contains the products customer i can buy in journey j (we also include $k = 0$, an additional "no-purchase product" in this set). We assemble the set of all observations $\mathcal{O} = \{(i, j, k)| \ i = 1, \ldots, I, \ j = 1 \ldots, J_i, \ k \in \mathcal{K}_j\}$, where each observation ("row" in our dataset) represents a product in a journey. We create a single training dataset using the clickstream data of the entire journey of each customer in the training data to estimate the benchmark models, which mimic the information seen by the proposed model.

To compute predictions in the test set (i.e., in journeys that have not been observed yet), we create a dataset that changes as information comes in. When making predictions after 5 steps, we use all the information in the journey available within the first 5 steps of the journey. To avoid selection bias and to be able to compare quantities across the different stages of the journey, we hold constant the set of journeys across the two test conditions: after query and after 5 steps (columns of Table 4). Specifically, for journeys shorter than 5 steps, we use the entire journey when making 5-step predictions.

For each observation, we create the binary outcome Y_{ijk}, which equals one if customer i purchased product k during journey j, and zero otherwise ($Y_{ij0} = 1$ if the customer ends the journey without a purchase); and a set of features X_{ijk} ("columns" in our dataset) that contain the information for each customer, journey, and product. Specifically, we include five types of features in X_{ijk}:

(1) the set of query variables for journey j (same as those in the query model),

(2) summary statistics of the attributes of all the products shown in the first page of journey j (same as those in the main model),

(3) the clicks and filters (during the focal journey) up to the moment when the prediction is made (capturing what the customer has been clicking so far),

(4) the queries, product attributes, clicks, filters, and purchases from past journeys (capturing the customer's past behavior), and

(5) the attributes of product k.

We now provide details about each of these sets of variables:

(1) We use the same set of query variables as in the main model. We encode all categorical variables as binary (leaving one level out to avoid multicollinearity).

(2) We use the same set of attributes as in the main model, with the exception that we encode categorical variables in full one-hot encoding, such that each level in a categorical variable has a corresponding binary feature. We summarize these features across all products shown on the first page of journey j, and compute the average, minimum, and maximum shown on the first page.

(3) We categorize 'clicks' in two primary ways. Firstly, at the product level, we represent using a binary feature whether the focal product k has been selected or not. In the training data, clicks throughout the entire journey are used to formulate this binary feature since the model undergoes a one-time training. In the test data, this feature is set to one if product k has been clicked on by that point in the journey. If the product remains unclicked, this feature corresponds to the percentage of products clicked in the training data; essentially, in the absence of the feature, we resort to the mean value from the training data.

Secondly, at the journey level, we aggregate the features of all clicked products within the focal journey, utilizing averages for continuous data and counts for binary data. Mirroring the process mentioned earlier, the training data summary is computed at the end of the journey, whereas the test data incorporates information accessible up to that specific step. Furthermore, we document the total count of clicked products within the focal journey. Filters are integrated in a similar fashion.

(4) We compute the average of variables (1) – (3) plus the attributes of purchased products and the number of past purchases, across all past journeys of customer i. In the training data, for focal journey j, these summaries are computed across journeys 1 through $j - 1$; whereas in the test data, we use the summary across all journeys of customer i in the training data. We also include the number of past journeys (such that a non-linear model can recreate counts).

(5) We include the features of product k as done in the proposed model, and we use a binary feature to distinguish between actual products and the No-Purchase product.

In sum, we generated a training dataset of 258,588 observations and 454 features. We train both binary classifiers, Random Forest (RF) and XGBoost, using a cross-entropy loss function (i.e., binary logistic). For the RF, we use honest splitting estimation, where the sample is split in two: one to construct the trees and another to evaluate the predictions. We use a sample fraction of 0.5, a number of variable tries per split of 41 ($\sqrt{\text{\#features}} + 20$), an honesty fraction of 0.5, and 2000 trees. For XGBoost, we use 100 rounds with a learning rate of 1 and a maximum depth of trees of 4.

After the models are trained, we compute predictions on the test data, $\widehat{pY}_{ijk}$, in multiple steps. First, we normalize the predictive scores from the benchmarks per journey, such that they sum to one by

$$\widehat{pY}_{ijk}^{\text{norm}} = \frac{\widehat{pY}_{ijk}}{\sum\limits_{k' \in \mathcal{K}(j)} \widehat{pY}_{ijk'}},$$

as these binary predictions are generated independently for all observations. Note that this normalization is not needed for the proposed model as the model provides a probability measure directly. The next steps apply to both benchmark models and our proposed model.

Second, for the incidence predictive task, we label a journey as a purchase if the normalized score for the no-purchase product is lower than 0.5, that is,

$$\widehat{Y}_{ij}^{\text{incidence}} = \mathbf{1}\left\{\widehat{pY}_{ij0}^{\text{norm}} \leqslant 0.5\right\}.$$

We compute balanced accuracy, precision, and recall from these predicted labels.

Third, for the product choice given purchase predictive task, we first compute choice given purchase scores per product by

$$\widehat{pY}_{ijk}^{\text{choice}} = \frac{\widehat{pY}_{ijk}^{\text{norm}}}{\sum\limits_{k' \in \mathcal{K}(j): k \neq 0} \widehat{pY}_{ijk'}^{\text{norm}}},$$

and label the predicted chosen alternative as the product with the maximum score per journey

$$\widehat{Y}_{ij}^{\text{choice}} = \underset{k \in \mathcal{K}(j)}{\arg\max} \left\{\widehat{pY}_{ijk}^{\text{choice}}\right\}.$$

We use the predicted labels $\widehat{Y}_{ij}^{\text{choice}}$ to compute the hit rate (percentage of journeys where predicted choice equals actual chosen product). In order to provide information on how the model predicts at the product level (what the models were trained for), we use $\widehat{pY}_{ijk}^{\text{choice}}$ to compute balanced accuracy by labeling as one the product with the highest score and computing the confusion matrix using the data at the journey-product level. Note that in such case, precision, recall, and balanced accuracy are all equal, as there is only one chosen product per journey (actual), and only one product is predicted to be chosen.

H.3 Measures of predictive performance

Due to the predominant occurrence of non-purchase outcomes in the majority of the journeys, we evaluate predictive ability in purchase incidence based on *balanced accuracy* (Brodersen et al. 2010) as it provides a more reliable measure of model performance when classes are imbalanced. It is calculated as the average of sensitivity/recall (true positive rate) and specificity (true negative rate) and therefore ranges from 0 to 1, where a value of 1 indicates perfect prediction performance. For product choice given incidence, we report hit rate (proportion of journeys that were correctly predicted) and balanced accuracy at the product level.

I Further details on the value of first-party data

Similarly to the analysis presented in Web Appendix H, we compute the choice probabilities for all models at each stage of the journey. We consistently employ all journeys across all stages, ensuring a constant set of journeys when conducting comparisons throughout the journey. For instance, when forecasting after 5 steps (or 2 steps), we consider the initial five (or two) steps for journeys with at least 5 (or 2) steps, while accommodating all available steps for journeys shorter than 5 (or 2) steps. This methodology enables us to assess the journey's value with a conservative lens, as performance on the held-out set would notably enhance if we were to observe a uniform 5 steps across all journeys.

We compute hit rates at the product level, analogous to the approach described in Section H. When exploring the ability of the model to predict what attributes the customer will choose, we compute the probabilities of choosing each level by aggregating the choice probabilities across all products with such a level. For categorical variables, we compute hit rates, and for continuous variables, we utilize the Root Mean Square Error (RMSE).

For example, let us consider a categorical attribute such as number of stops. For each level — Non-stop, One stop, and 2+ stops — we compute the probability that a customer, conditional on making a purchase, will opt for a specific stop level. This is done by aggregating the choice probabilities associated with the 'stop' attribute. For instance, the probability that a customer will select a non-stop flight corresponds to the cumulative choice probabilities of all non-stop flights. Subsequently, the predicted number of stops is identified as the level with the highest choice probability. We then contrast these predicted labels with the actual labels to compute hit rates, which represent the proportion of journeys where we accurately predict the number of stops for the chosen flight. A similar methodology is applied when considering airline alliances, which is also categorical.

For a continuous attribute such as price, we first calculate the square errors between the price of each alternative and the price of the purchased alternative, and then compute the weighted average of those square errors (by journey) using the purchase probabilities as weights, by

$$MSE_{ij}^{\text{Price}} = \sum_{k \in \mathcal{K}(j)} p(y_{ij}^{p} = k|\text{Data}_{ijt}) \cdot \left(\text{Price}_{ijk} - \text{Price}_{ijk*}\right)^2,$$

where $p(y_{ij}^{p} = k|\text{Data}_{ijt})$ are the purchase probabilities and k^* is the true purchased alternative. First, note that the square errors are independent from the predictions, but the weighted average is not. Second, note that if the model predicts with probability one on alternatives with the same price as the purchased one, then this expectation is zero. Finally, we average those expected square errors and compute the square root

$$RMSE^{\text{Price}} = \sqrt{\frac{1}{J^{oos}} \sum_{ij} MSE_{ij}^{\text{Price}}},$$

where J^{oos} is the number of held-out journeys. We follow the same procedure for the length of the trip. We compute these scores on the normalized prices and lengths to weigh all journeys equally and to avoid searches with more expensive and longer destinations to dominate the score.

J Further details on the retargeting analysis

We simulate outcomes for a single ad impression of a retargeted ad for each customer where we vary the product featured in the impression. For each journey with a minimum of two clicks in the holdout sample, we select a product to be displayed in a single ad impression. Using our model and the data up to the current step in the journey, we use the posterior inferences on $\boldsymbol{\beta}_{ij}$ to compute the posterior expected utility for each product in the journey and select the product with the maximum expected utility. We compare this product chosen against two benchmarks. First, we select the product that is expected to be the most popular a priori, without any clickstream and personalized data. Since product availability is quite sparse across journeys as a specific product is unlikely to appear across multiple journeys, we use the population average preferences from the `Only purchase` model estimated in Section 5, and compute expected utility under the population preferences to choose the product with highest expected utility across the population. Second, we select the product that has been clicked last on each journey. In those journeys where no product has been clicked on we select the most popular product described in the previous benchmark.

For each retargeting ad, we simulate click-through-rates using the estimates of a model with full information of the journey, that is, those using all data from those journeys, including clicks after the second step and purchases (which has different estimates than those used for setting the policy). We simulate each ad impression click occasion as a univariate probit model with utilities

$$u_{ij}^{r}(k^{*}) = \beta^{0r} + \boldsymbol{\beta}_{ij}^{x}{}'\mathbf{x}_{ijk^{*}} + \varepsilon_{ij}^{r}$$

where k^{*} is the chosen product for each retargeting method, β^{0r} is a retargeting ad intercept and $\varepsilon_{ij}^{r} \sim \mathcal{N}(0, \sigma_r^2)$ is the unobserved component of utility. Since both the intercept β^{0r} and the error term variance σ_r^2 are not parameters in our model, we set them to different values. The intercept β^{0r} is set such that the mean CTR across all products and journeys for a retargeted ad equals 2% across journeys (the average retargeting CTR without content personalization reported in the industry is 0.7% (Signifi Media 2020), and highly personalized ads have 3x higher CTRs compared to non-personalized ads (Bleier and Eisenbeiss 2015)). We set the variance of the error term to $\sigma_r^2 = 1.25$ and vary it for robustness.

We show in Table J.7 the predicted CTR across retargeting methods to choose the personalized product and choices of σ_r^2. We note that even though the base CTR is sensitive to the choice of σ_r^2, the relative improvement is robust.

Table J.7: Retargeting: Predicted CTR when the firm utilizes insights from the model to formulate a retargeting offer. The baseline models presume that the retargeting offer features the product that was most recently clicked or the most popular product.

Model	**Click-through rates**		
	σ =1.00	σ =1.25	σ =1.50
Baseline 1 (Most popular)	0.108	0.054	0.027
Baseline 2 (Last clicked)	0.111	0.056	0.028
% increase vs. most popular	+3.05%	+3.65%	+3.93%
Highest preference based on our model	0.143	0.074	0.037
% increase vs. most popular	+31.94%	+36.27%	+39.16%
% increase vs. last clicked	+28.04%	+31.47%	+33.90%