9 Web Appendix A:

9.1 Further evidence for section 3

		Motorola Droid RAZR MAXX HD	Apple iPhone 5	HTC One	Nokia Lumia 920	Samsung Galaxy Note II
	Colors Available	Click to Buy	Click to Buy	Click to Buy	Click to Buy	Click to Buy
Price	Price with 2-year Contract	$199.99	$249.99	$199.99	$99.99	$249.99
Battery	Battery Type	Lithium-ion Polymer	Lithium-ion	Lithium-polymer	Lithium-polymer	Lithium-ion
	Standby Time	Up to 16 days	Up to 9 days	Up to 20 days	Up to 13 days	Up to 19.5 days
	Talk Time	Up to 32 hours	Up to 8 hours	Up to 18 hours	Up to 7.5 hours	Up to 21 hours
Memory	Internal Memory	32GB	32GB	32GB	32GB	16GB
	Memory Card	microSD up to 32GB	No	No	No	microSD up to 64GB
	RAM	1GB	1GB	2GB	1GB	2GB
Wireless Capabilities	Band and Mode	|EDGE|GSM|Quad-band GSM|3G|UMTS|HSPA+|HSDPA |CDMA|4G LTE|	|EDGE|GSM|Quad-band GSM|3G|UMTS|Quad-band UMTS|HSPA+|HSDPA|CDMA|4G LTE|Dual-band LTE|	|EDGE|GSM|Quad-band GSM|GPRS|3G|UMTS|HSDPA |CDMA|4G|4G LTE|Dual-band LTE|	|EDGE|GSM|Quad-band GSM|3G|Quad-band UMTS|HSPA+|HSDPA|4G|4G LTE|Dual-band LTE|	|GSM|3G|UMTS|HSPA+|HSDPA |CDMA|4G LTE|Dual-band LTE|Quad-band|
	Bluetooth	4.0	4.0	4.0	3.0 + HS	4.0
	Built-In GPS	Yes	Yes	Yes	Yes	Yes
	WiFi	802.11 a/b/g/n	802.11 a/b/g/n	802.11 a/b/g/n/ac	802.11 a/b/g/n	802.11 a/b/g/n
Camera	Camera Resolution	8.0MP	8.0MP	HTC UltraPixel 4MP Camera	8.7MP	8.0MP
	Records Video	1080p Full HD	1080p Full HD	1080p Full HD	1080p Full HD	1080p Full HD
	Secondary Camera	1.3MP	1.2MP	2.1MP	1.2MP	1.9MP
Display	Pixel Density	313 PPI	326 PPI	468 PPI	332 PPI	267 PPI
	Screen Resolution	1280 x 720	1136 x 640	1920 x 1080	1280 x 768	1280 x 720
	Screen Size	4.7 inches	4.0 inches	4.7 inches	4.5 inches	5.5 inches
Size	Product Dimensions	5.19" x 2.67" x 0.37"	4.87" x 2.31" x 0.30"	5.41" x 2.69" x 0.37"	5.13" x 2.79" x 0.42"	5.95" x 3.17" x 0.37"
	Product Weight	5.54 oz.	3.95 oz.	5.04 oz.	6.53 oz.	6.44 oz.
Operating System (OS) and Other Features	CPU	1.5 GHz Dual Core	Not Specified by manufacturer	1.7 GHz Quad Core	1.5 GHz Dual Core	1.6 GHz Quad Core
	Media - Audio	AAC, AAC+, AMR-NB, AMR-WB, eAAC+, MIDI, MP3, OGG, WMA v9, WMA v10	AAC, HE-AAC, MP3, MP3 VBR, Audible, Apple Lossless, AIFF, and WAV	AAC, AMR, OGG, M4A, MID, MP3, WAV, WMA	MP3, QCELP, AMR-WB, AMR-NB, WMA 10 Pro, WMA 9, G.711, AAC-LC, AAC+/HEAAC, eAAC+/HEAACv2, ASF, MP4, AAC, M4A, 3GP, 3GZ	MP3, OGG, WMA, AAC, ACC+, eAAC+, AMR-NB, AMR-WB, MIDI, WAV, AC-3, Flac
	Mobile Operating System	Android 4.1 Jelly Bean	iOS 6	Android 4.1 Jelly Bean	Windows Phone 8	Android 4.1 Jelly Bean
	NFC	Yes	No	Yes	Yes	Yes
	QWERTY Keyboard	No	No	No	No	No
	Supported Email	POP3, IMAP, Push email (Exchange, Gmail)	POP3, IMAP, Push email (Exchange, Gmail)	POP3, IMAP, Push email (Exchange, Gmail, Hotmail)	SMTP, POP3, IMAP4, Push email (Exchange, Gmail, Hotmail)	POP3, IMAP, Push email (Exchange, Gmail)
	Voice Dial	Yes	Yes	Yes	Yes	Yes

Figure A-1: Experiment design under the medium complexity condition

		HTC One	Nokia Lumia 920	Motorola Droid RAZR MAXX HD	Samsung Galaxy Note II	Apple iPhone 5
	Colors Available	Click to Buy	Click to Buy	Click to Buy	Click to Buy	Click to Buy
Price	Price with 2-year Contract	$199.99	$99.99	$199.99	$249.99	$249.99
Operating System (OS) and Other Features	CPU	1.7 GHz Quad Core	1.5 GHz Dual Core	1.5 GHz Dual Core	1.6 GHz Quad Core	Not Specified by manufacturer
	Media - Audio	AAC, AMR, OGG, M4A, MID, MP3, WAV, WMA	MP3, QCELP, AMR-WB, AMR-NB, WMA 10 Pro, WMA 9, G.711, AAC-LC, AAC+/HEAAC, eAAC+/HEAACv2, ASF, MP4, AAC, M4A, 3GP, 3GZ	AAC, AAC+, AMR-NB, AMR-WB, eAAC+, MIDI, MP3, OGG, WMA v9, WMA v10	MP3, OGG, WMA, AAC, ACC+, eAAC+, AMR-NB, AMR-WB, MIDI, WAV, AC-3, Flac	AAC, HE-AAC, MP3, MP3 VBR, Audible, Apple Lossless, AIFF, and WAV
	Media - Video	3GP, 3G2, MP4, WMV, AVI	H.264 / AVC, MPEG-4, VC-1, Windows video, H.263, WMV, AVI, 3GP, 3G2, M4V, MOV	MPEG-4, H.263, H.264, VC-1, VP8	MPEG4, H.263, H.264, VC-1, DivX, WMV, VP8, 3GP(MP4), AVI, FLV, MKV, WebM	h.264 / AVC, Motion JPEG, MPEG-4, Quicktime
	MMS	Yes	Yes	Yes	Yes	Yes
	Mobile Operating System	Android 4.1 Jelly Bean	Windows Phone 8	Android 4.1 Jelly Bean	Android 4.1 Jelly Bean	iOS 6
	NFC	Yes	Yes	Yes	Yes	No
	OS Support	Windows (7, Vista, XP), Mac OS	Windows (7, Vista, XP), Mac OS	Windows (7, Vista, XP), Mac OS	Windows (7, Vista, XP), Mac OS	Windows (7, Vista, XP), Mac OS (OS X)
	Phone Style	Bar phone	Bar phone	Bar phone	Bar phone	Bar phone
	QWERTY Keyboard	No	No	No	No	No
	Released (US)	4/19/2013	11/9/2012	10/18/2012	11/9/2012	9/21/2012
	Speakers	Stereo	Mono	Mono	Mono	Mono
	Supported Email	POP3, IMAP, Push email (Exchange, Gmail, Hotmail)	SMTP, POP3, IMAP4, Push email (Exchange, Gmail, Hotmail)	POP3, IMAP, Push email (Exchange, Gmail)	POP3, IMAP, Push email (Exchange, Gmail)	POP3, IMAP, Push email (Exchange, Gmail)
	TV-out	MHL / HDMI (micro-USB), Wireless with DLNA	Wireless with DLNA	Micro-HDMI, Wireless with DLNA	MHL / HDMI (micro-USB)	Lightning Digital AV Adapter, Lightning to VGA Adapter, Wirelessly
	Voice Dial	Yes	Yes	Yes	Yes	Yes
	Warranty (Labor/Parts)	1 Year	1 Year	1 Year	1 Year	1 Year
Memory	Internal Memory	32GB	32GB	32GB	16GB	32GB
	Memory Card	No	No	microSD up to 32GB	microSD up to 64GB	No
	RAM	2GB	1GB	1GB	2GB	1GB
Size	Product Dimensions	5.41" x 2.69" x 0.37"	5.13" x 2.79" x 0.42"	5.19" x 2.67" x 0.37"	5.95" x 3.17" x 0.37"	4.87" x 2.31" x 0.30"
	Product Weight	5.04 oz.	6.53 oz.	5.54 oz.	6.44 oz.	3.95 oz.
Wireless Capabilities	Band and Mode	|EDGE|GSM|Quad-band GSM|GPRS|3G|UMTS|HSDPA |CDMA|4G|4G LTE|Dual-band LTE|	|EDGE|GSM|Quad-band GSM|3G|Quad-band UMTS|HSPA+|HSDPA|4G|4G LTE|Dual-band LTE|	|EDGE|GSM|Quad-band GSM|3G|UMTS|HSPA+|HSDPA |CDMA|4G LTE|	|GSM|3G|UMTS|HSPA+|HSDPA |CDMA|4G LTE|Dual-band LTE|Quad-band|	|EDGE|GSM|Quad-band GSM|3G|UMTS|Quad-band UMTS|HSPA+|HSDPA|CDMA|4G LTE|Dual-band LTE|
	Bluetooth	4.0	3.0 + HS	4.0	4.0	4.0
	Built-In GPS	Yes	Yes	Yes	Yes	Yes
	WiFi	802.11 a/b/g/n/ac	802.11 a/b/g/n	802.11 a/b/g/n	802.11 a/b/g/n	802.11 a/b/g/n
Camera	Camera Resolution	HTC UltraPixel 4MP Camera	8.7MP	8.0MP	8.0MP	8.0MP
	Records Video	1080p Full HD	1080p Full HD	1080p Full HD	1080p Full HD	1080p Full HD
	Secondary Camera	2.1MP	1.2MP	1.3MP	1.9MP	1.2MP
Battery	Battery Type	Lithium-polymer	Lithium-polymer	Lithium-ion Polymer	Lithium-ion	Lithium-ion
	Standby Time	Up to 20 days	Up to 13 days	Up to 16 days	Up to 19.5 days	Up to 9 days
	Talk Time	Up to 18 hours	Up to 7.5 hours	Up to 32 hours	Up to 21 hours	Up to 8 hours
Display	Pixel Density	468 PPI	332 PPI	313 PPI	267 PPI	326 PPI
	Screen Resolution	1920 x 1080	1280 x 768	1280 x 720	1280 x 720	1136 x 640
	Screen Size	4.7 inches	4.5 inches	4.7 inches	5.5 inches	4.0 inches
	Screen Type	LCD (Active, Color, Backlit)	LCD (Active, Color, Backlit)	OLED (Active, Color, Backlit)	OLED (Active, Color, Backlit)	LCD (Active, Color, Backlit)
	Touch Screen	Yes	Yes	Yes	Yes	Yes

Figure A-2: Experiment design under the high complexity condition

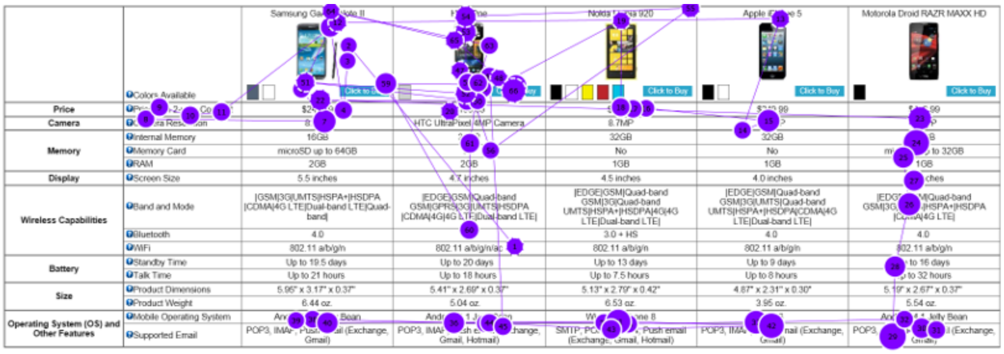

Figure A-3: Illustrating eye-fixations

10 Web Appendix B: Robustness checks

10.1 Model without learning: Weitzman (1979)

To compare our results to a model of sequential search without learning, in this section we estimate the Weitzman (1979) model using our data. Note that this model can explain the extensive margin of search (which brands were searched), but not the intensive margin of search (how many fixations were performed on each searched brand). Thus, we collapse all fixations of a consumer at the brand level and try to understand consumer search and purchase decisions. This modification drastically reduces the number of observations and limits the model's ability to fully account for all search decisions consumers make in our study, limitation which our main model presented in Section 5 does not suffer from.

Consider the decision of a consumer who is searching among the five brands available, $J = \{1, \ldots, 5\}$. The utility of consumer i from purchasing brand j is given by

$$\begin{aligned} u_{ij} &= v_{ij} + \epsilon_{ij} \\ &= Brand_j + \theta X_{ij} + \eta_{ij} + \epsilon_{ij}. \end{aligned} \tag{B1}$$

In this model, it is assumed that consumers know the distribution of utilities for each brand $N(v_{ij}, \sigma^2)$, but are searching to reveal the match value ϵ_{ij}. In other words, consumers know the true utility distribution (there is no learning), and all uncertainty about an alternative is revealed in one search. The value v_{ij} represents any information consumers have about the brand before starting their search. We model this information as a function of brand intercepts, an unobserved shock η_{ij} (from the researcher perspective) and potentially additional controls, such as the consumer's prior ownership and familiarity with each brand.[33] This specification allows for the possibility that prior information affects consumers' expected benefit from search in a setting without learning. The match value ϵ_{ij} is revealed through fixations performed on a brand (unobserved both before and after search by the

[33] The utility specification in equation B1 differs from the one in our main model (equation 5), because of the different learning assumptions made in the two models. Also note that we do not take the third measure of prior information, prior experience, into account in this model since it does not vary across brands and consumers cannot search across attributes in the Weitzman model.

researcher, observed after search by the consumer). Search is costly. We assume that search costs are given $c_i = exp(\kappa_i)$.

According to Weitzman (1979), optimal search follows an index policy: consumers attach an index, called a reservation utility, to each brand, and search in decreasing order of this index (*selection rule*). This reservation utility z_{ij} is the unique solution to

$$c_{ij} = \int_{z_{ij}}^{\infty} (u - z_{ij}) dF_{ij}(u), \tag{B2}$$

where $F_{ij}(u)$ coincides in our case with the utility distribution $N(v_{ij}, \sigma^2)$. To solve for z_{ij} from equation B2 above, see Kim et al. (2010). When the best option observed so far exceeds the index of the brand under consideration, search stops, otherwise the consumer continues searching the brand with the next highest index (*search rule*). Once search ceases, the consumer chooses to purchase the brand with the highest observed utility among those searched (*choice rule*). We assume there is no outside option, since consumers in our experiment choose among the five options available. Nokia serves as a reference brand, as it did in the main model.

More formally, to construct the likelihood function and estimate this model, we use the search rules described above to impose restrictions on the parameters of interest that made it optimal for the consumer to search and purchase the options we observe. To describe these restrictions, suppose that we observe a consumer searching a number s of the total J options available, and that she chose to purchase option j. With a slight abuse of notation, order brands by their reservation utilities and let n denote the brand with the nth largest reservation utility. Since consumers search brands in a decreasing order of their reservation utilities, according to the selection rule, it must be that

$$z_{in} \geq \max_{k=n+1}^{J} z_{ik} \qquad \forall n \in \{1, \ldots, J-1\}. \tag{B3}$$

Since all consumers in our data search at least once, consistent with prior work (e.g., Honka 2014, Honka and Chintagunta 2016), we assume that the first search is free.

According to the stopping rule, for all brands that were searched, it must be that

$$z_{in} \geq \max_{k=0}^{n-1} u_{ik} \qquad \forall n \in \{1, \ldots, s\}, \tag{B4}$$

while for all brands that were not searched, it must be that

$$z_{im} \leq \max_{k=0}^{s} u_{ik} \qquad \forall m \in \{s+1, \ldots, J\}. \tag{B5}$$

Finally, consistent with the choice rule, if the consumer chooses j, her utility from this choice exceeds that of all searched brands, i.e.,

$$u_{ij} = \max_{k=0}^{s} \; u_{ik} \qquad \forall j \in \{0, 1, \ldots, s\}. \tag{B6}$$

If consumers search using the rules described above, then they make search and purchase decisions jointly. Thus, the probability of observing a certain outcome in the data for consumer i is characterized by the joint probability of equations (B3)-(B6) holding. This probability is given by

$$L_i = Pr(\text{Selection rule}_i, \text{ Search rule}_i, \text{ Choice rule}_i). \tag{B7}$$

Because consumers make these decisions jointly, the likelihood function does not have a closed-form solution. We use a simulated maximum likelihood (SMLE) approach to estimate the parameters of the model following McFadden (1989), Honka (2014), Honka and Chintagunta (2016), Ursu (2018), Ursu et al. (2020). This approach involves the following steps:

1. Make $d = \{1, \ldots, D\}$ draws of η_{ij} and ϵ_{ij} for each consumer-brand combination and calculate utility u_{ij}^d.
2. Compute z_j^d.
3. Calculate the following expressions for each draw d:
   (a) $\nu_1^d = z_{in}^d - \max_{k=n+1}^J z_{ik}^d \quad \forall n \in \{1, \ldots, J-1\}$
   (b) $\nu_2^d = z_{in}^d - \max_{k=0}^{n-1} u_{ik}^d \quad \forall n \in \{1, \ldots, s\}$
   (c) $\nu_3^d = \max_{k=0}^s u_{ik}^d - z_{im}^d \quad \forall m \in \{s+1, \ldots, J\}$
   (d) $\nu_4^d = u_{ij}^d - \max_{k=0}^s u_{ik}^d \quad \forall j \in \{0, 1, \ldots, s\}$
4. Compute $V^d = \frac{1}{1+M^d}$ for each draw d, where

$$M^d = \sum_{k=1}^{4} e^{-\nu_k^d/\rho}, \tag{B8}$$

   where ρ is scaling parameter, which we set to 1/10.
5. The average of V^d over the D draws and over consumers and brands gives the simulated likelihood function.

We estimate the model with $D = 1{,}000$ utility error terms and repeat the estimation 50 times, reporting average results in the table below.

Our results can be found in Table B-1 below. We present results using two definitions of a brand search. In the first case, we assume any brand that a consumer fixated on at least once was searched (Model 1). In the second case, we define as searched only those brands that a consumer fixated on at least 10 times (Models 2 and 3). The second case considers the situation where consumers require a certain minimum number of fixations to reveal ϵ_{ij} that we do not observe.

Table B-1: Weitzman Model Estimation Results

	Model 1: Search if fixation> 0	Model 2: Search if fixation≥ 10	Model 3: Search if fixation≥ 10
	(1)	(2)	(3)
Utility			
Samsung	0.2059***	0.2889***	0.1704+
	(0.0579)	(0.0583)	(0.0945)
Apple	0.3154***	0.3530***	0.1924*
	(0.0589)	(0.0651)	(0.0960)
HTC	0.1378*	0.1587**	0.1134+
	(0.0560)	(0.0569)	(0.0885)
Motorola	−0.0808	−0.0345	−0.0821
	(0.0632)	(0.0600)	(0.1012)
Nokia	-	-	-
Prior Ownership			0.2237*
			(0.0966)
Familiarity			0.3647+
			(0.2017)
Search Cost (exp)			
Constant	−10.9494***	−5.9909***	−6.1162***
	(0.4055)	(0.2409)	(0.2317)
LL	-2,278	-2,331	-2,218
AIC	4,567	4,471	4,449
BIC	4,594	4,498	4,487
Observations	1,710	1,625	1,625

Standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Estimation results for the Weitzman (1979) model, using 1000 utility error term draws. We repeated the estimation 50 times, reporting here the average results. For consistency, in Models 2 and 3, because only brands on which consumers fixate at least 10 times are considered as searched, we drop consumers who choose brands after fewer than 10 fixations, as well as those with no searches under this definition (17 consumers were dropped).

Across all models, we find that Apple and Samsung are the most valuable brands and that prior information has a positive effect on consumer utility. Also, the third model replicates our previous results that prior ownership and familiarity increase utility, and accounting for these measures of prior information lowers brand estimates and improves fit.

These results fail to account for the intensity of search (number of fixations per brand), the sequence of search decisions, as well as the switches from and revisits of a brand. Therefore, compared to our model, here we report relatively low mean search cost estimates. This occurs because 95% of consumers search all brands and 61% of consumers make at least 10 fixations on all brands. Thus, the Weitzman (1979) model interprets the large number of searches as an indication of low search costs. In contrast, our model more accurately captures the fact that consumers do not process all relevant brand information at once, obtaining relatively more information about some brands, while ignoring others. This reveals that consumers have relatively higher search costs. In addition, this model cannot capture the effect of prior experience since this model cannot capture search across brands and attributes, but only across brands. Finally, we find that search costs when accounting for prior information (model 3) are lower than when this information is omitted (model 2). This finding differs from the one in our main specification for at least two reasons: (i) without learning, the precision of the signals obtained through learning cannot adjust to rationalize the same number of searches, and (ii) model 3 cannot account for the effect of prior experience on search costs. These reasons, as well as the patterns supporting consumer learning which we presented in Section 4, suggest that a model of search with learning is better suited to model search behavior in our data.

10.2 Alternative specifications of the role of prior information

In this section, we present results obtained by estimating the model presented in Section 5 under alternative specifications of the relation between prior information and model primitives. In Table B-2 we show results where we test almost all possible combinations of the effect of prior information that are feasible in our data. As can be seen, all alternatives lead to a worse fit than the model we presented as our main specification. In addition, most of these results show the same patterns in relation to the results with no prior information we presented in Table 5: brand intercepts are smaller and search costs are larger than if no prior information was included in the model.

Table B-2: Estimation Robustness Checks

Prior mean	*PO*	*Fam*	*PO & Fam*	*Fam*	*PO & Fam*
Prior variance	*Fam*	*PO*		*PO*	
Signal variance	*Exp*	*Exp*	*Exp*		
Search cost				*Exp*	*Exp*
	(1)	(2)	(3)	(4)	(5)
Brand Value					
Samsung	0.4475***	0.4645***	0.4917***	0.3238***	0.3218***
	(0.0883)	(0.0511)	(0.0749)	(0.0843)	(0.0842)
Apple	0.4680***	0.5405**	0.5272**	0.3544**	0.3527***
	(0.0811)	(0.1770)	(0.1752)	(0.1330)	(0.1068)
HTC	0.2120***	0.2763***	0.2386***	0.1695+	0.1686
	(0.0686)	(0.0414)	(0.0281)	(0.1172)	(0.1526)
Motorola	0.1636+	0.1778	0.1857	0.1274	0.1271
	(0.0902)	(0.2090)	(0.1492)	(0.2266)	(0.2128)
Nokia	-	-	-	-	-
Signal variance					
Constant	0.4176***	0.5561***	0.5324***	0.2059***	0.2025***
	(0.0304)	(0.0583)	(0.1589)	(0.0578)	(0.0477)
Search Costs (exp)					
Constant	−1.5680***	−1.9585***	−1.9249***	−1.0308***	−1.0247***
	(0.0312)	(0.0511)	(0.0749)	(0.0506)	(0.0285)
Prior Information					
Prior Ownership	0.3791***	0.3965+	0.3929+	0.3599*	0.3599**
	(0.0000)	(0.2709)	(0.2684)	(0.1627)	(0.1234)
Familiarity	0.6510***	0.6337***	0.6437***	0.5112***	0.5101***
	(0.0744)	(0.0583)	(0.1597)	(0.0083)	(0.0785)
Prior Experience	0.0572***	0.0503	0.0517	−0.6268***	−0.6231***
	(0.0001)	(0.2486)	(0.1440)	(0.0280)	(0.0708)
LL	−86,082	−86,060	−86,063	−86,140	−86,141
AIC	172,183	172,138	172,144	172,297	172,300
BIC	172,259	172,214	172,220	172,373	172,377
Observations	36,170	36,170	36,170	36,170	36,170

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Estimation results under alternative specifications of the relation between prior information and model primitives. The results are averaged over 10 estimations using different starting seeds, with 50 signal draws each.

10.3 Additional simulation and estimation results

Table B-3: Monte Carlo simulation results - Latent class model

	True	*Segment 1*	*Segment 2*
	(1)	(2)	(3)
Prior mean			
Prior Ownership	{2.00, 1.00}	2.0619	1.0682
		(0.4788)	(0.6371)
Prior variance			
Familiarity	{1.00, 0.50}	1.0538	0.5857
		(0.6015)	(0.6163)
Signal mean			
Brand Value	{1.00, 0.50}	1.2001	0.5110
		(0.4643)	(0.1166)
Signal variance			
Constant	{1.00, 2.00}	1.0877	1.9889
		(0.5074)	(0.3988)
Search cost (exp)			
Constant	{-1.00, -2.00}	-0.9564	-2.0958
		(0.2399)	(0.5563)
Prior Experience	{-2.00, 1.00}	-1.9477	1.1112
		(0.5977)	(0.4920)
Segment 1 probability $\pi_1 = exp(\rho)/(1+exp(\rho))$			
ρ	{0.40}	0.4524	
		(0.5022)	
Observations	30,089		
LL	−109,773		

Notes: Data is simulated for 1000 consumers, 5 brands and 7 attributes and the reported estimation results are obtained after averaging across estimation results from 200 different seeds, with 100 signal draws each. The standard deviation of the mean estimate across these simulations is reported in parentheses. The number of observations varies slightly across seeds, with the number reported above representing the average number of observations.

Table B-4: Additional Monte Carlo simulation results with an alternative true parameter vector

	True	*With Prior Information* Estimate	*Without Prior Information* Estimate
	(1)	(2)	(3)
Prior mean			
Prior Ownership	1.50	1.4867	-
		(0.8786)	
Prior variance			
Familiarity (inverse)	0.50	0.5228	-
		(0.3002)	
Signal mean			
Brand Value	0.80	0.7750	0.9112
		(0.0826)	(0.1649)
Signal variance			
Constant	1.00	0.9013	2.4865
		(0.5280)	(1.4859)
Search cost (exp)			
Constant	-1.30	-1.2927	-1.7732
		(0.1094)	(0.2339)
Prior Experience	-1.00	-0.8961	-
		(0.2783)	
Observations		31,135	31,135
LL		−111,709	−112,139

Notes: Data is simulated for 1000 consumers, 5 brands and 7 attributes and the reported estimation results are obtained after averaging across estimation results from 50 different seeds, with 100 signal draws each. The standard deviation of the mean estimate across these simulations is reported in parentheses. The number of observations varies slightly across seeds, with the number reported above representing the average number of observations.

Table B-5: Estimation results - Latent class model

	Segment 1 (1)	*Segment 2* (2)
Prior Mean		
Prior ownership	0.3609^{*}	0.3102^{+}
	(0.1484)	(0.2182)
Prior Variance		
Familiarity (inverse)	0.5482^{***}	0.5270^{***}
	(0.1196)	(0.0107)
Signal Mean		
Samsung	0.3077^{***}	0.3102^{**}
	(0.0395)	(0.1039)
Apple	0.3634^{+}	0.3489^{+}
	(0.2342)	(0.2223)
HTC	0.1743	0.1656
	(0.3641)	(0.3403)
Motorola	0.1341	0.1205
	(0.3316)	(0.2749)
Nokia	-	-
Signal Variance		
Constant	0.2348^{**}	0.1989^{+}
	(0.0896)	(0.1217)
Search Cost (exp)		
Constant	-0.3413^{***}	-0.5535^{***}
	(0.0202)	(0.0289)
Prior Experience	-0.6620^{***}	0.1010^{***}
	(0.0344)	(0.0072)
Segment 1 probability $\pi_1 = exp(\rho)/(1+exp(\rho))$		
ρ	0.1344	
	(0.1251)	
LL	−85,676	
AIC	171,390	
BIC	171,552	
Observations	36,170	

$^{+}$ ($p < 0.10$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$)

Notes: The results are averaged over 10 estimations using different starting seeds, with 50 signal draws each.

Table B-6: Estimation results - Alternative assumption: brand intercepts sum to 0

	With Prior Information (1)
Prior Mean	
Prior ownership	0.3772^{***}
	(0.0249)
Prior Variance	
Familiarity (inverse)	0.5967^{***}
	(0.0428)
Signal Mean	
Samsung	0.2125^{***}
	(0.0456)
Apple	0.3323^{***}
	(0.0667)
HTC	0.1146^{+}
	(0.0648)
Motorola	0.0467
	(0.0453)
Nokia	-
Signal Variance	
Constant	0.3256^{***}
	(0.0194)
Search Cost (exp)	
Constant	-0.3729^{***}
	(0.0023)
Prior Experience	-0.6422^{***}
	(0.0080)
LL	−85,763
AIC	171,544
BIC	171,621
Observations	36,170

$^{+}$ ($p < 0.10$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$)

Notes: The results are averaged over 10 estimations using different starting seeds, with 50 signal draws each.

11 Web Appendix C: Estimation sample

The main data set we described in Section 3.2 consists of 342 consumers and 78,617 observations (78,275 eye-fixations and 342 individual consumer brand choices). The average (median) number of fixations in these data totaled 229 (186). Due to the very large number of fixations that some consumers made in our data, estimating our model on the full data proved computationally very challenging.

Therefore, we focused our estimation on those consumers who made at most 300 fixations in our data. Our final estimation sample contains 253 consumers (74% of the original set of consumers) and 36,170 observations (35,917 fixations and 253 individual consumer brand choices). The average (median) number of fixations in these data totaled 142 (131).

12 Web Appendix D: Data generating process

In this Appendix, we describe the process our model assumes generated the data. The same process will be used to generate the data in our Monte Carlo simulations. This process is also visible in our estimation and simulation codes.

We start by describing the structure of the data. This is illustrated through an example in Figure D-1 below. A consumer $i \in \{1,\dots,I\}$ searches among $j \in J$ brands with $l \in L$ attributes each (a total of $J \times L$ options) at every time $t \leq T_i$ before stopping her search at T_i. Therefore, the data on consumers' choices will be structured as follows: for every consumer, there will be T_i rows and $J \times L + 1$ columns. In rows 1 through $T_i - 1$, an indicator will specify which of the $J \times L$ options (with a 1 entering the respective column) the consumer searched. For example, if we ordered columns such that column 1 indicates brand 1, attribute 1, column 2 indicates brand 2, attribute 1, etc, then in Figure D-1 we see that the consumer started her search by looking at brand 3 attribute 1. She then continued searching at $t = 2$ by looking at the same brand and attribute, while at $t = 3$ she switched to looking at brand 2 attribute 1, and so on. In row T_i and column $J \times L + 1$, a number $1 : J$ will indicate which of the brands was chosen by the consumer when she ceased her search. In our example, the consumer decides to stop and choose brand 3 at T_i, which ends her search.

Data on consumer i's choices							
time/brand-attr.	1	2	3	...	...	jxl	jxl+1
1	0	0	1	0	0	0	0
2	0	0	1	0	0	0	0
3	0	1	0	0	0	0	0
...	1	0	0	0	0	0	0
...	0	0	0	0	0	1	0
Ti-1	0	0	1	0	0	0	0
Ti	0	0	0	0	0	0	3

Figure D-1: Data structure: Example

In our estimation sample, there are 253 consumers, 5 brands, and 7 attributes. We observe consumers' choices (data with the same structure as in Figure D-1) and their prior information: level of prior brand ownership (0 or 1 for each consumer-brand combination), familiarity (scale from 1 to 7 for every consumer-brand combination) and attribute experience (scale from 0 to 1 for every consumer-attribute combination). Using these data we will try to recover consumer preferences and search costs. In our Monte Carlo simulation, there are 2,000 consumers, 5 brands, and 7 attributes. We will generate data on prior beliefs and choices for a given set of parameters, and then try to recover those parameters as we would in our estimation (take data on prior beliefs and choices and estimate parameters). In what follows, we first describe how we generate data in our Monte Carlo simulation. Then, in Appendix 13, we describe our estimation routine to recover parameters based on these data (complementing the discussion from Section 5.4).

There are two sets of data we need to generate: (1) data on consumer prior beliefs and (2) data on searches and choices. When generating data on prior beliefs, our goal is to create variables that are as close as possible to our data and ensure enough random variation. For prior ownership, each consumer will have a 60% chance of owning one of the brands. We implement this step by making a draw for every consumer from the uniform distribution from 0 to 1. If this draw is smaller than the value 0.6, then the consumer owns a brand, the identity of which is randomly drawn (random sample from 1 to J without replacement). For familiarity (prior experience), each consumer-brand combination (consumer-attribute combination) is populated with a draw from 1 to 10 (random sample from 1 to 10 with replacement), value which is then divided by 10 to ensure that variables in the model are on the same scale (approximately 0 to 1), which mirrors the scale of these variables in our data.

A set of parameters dictate consumer choices (same notation as in paper): μ determines brand intercepts (signal mean), β determines the effect of prior ownership, γ determines the familiarity effect, while θ affects the signal variance. Also, κ and ω determine the mean search cost and the effect of prior experience on search costs. In the simulation, we generate data using the parameters we display in Table 4. In general, for any set of parameters, we can generate consumer searches and choices, as we describe below.

To generate data on searches and choices, we start by constructing the variables of interest that affect such decisions: the prior mean, prior variance, signal mean, signal variance, and search costs. For this, we will use the parameters above and the equations we introduced in the paper, which we reproduce here. For the prior mean, we combine data on prior ownership and the β parameter to form:

$$\mu_{ij0} = \lambda_i + \beta_i PriorOwnership_{ij}, \tag{D1}$$

for every consumer-brand combination. As mentioned in the paper, λ_i is given by the average of the brand intercepts, so $\lambda_i = \frac{\sum_{j=1}^{J} \mu_{ij}}{J}$. In Section 5.5, we impose the restriction that brand intercepts sum to 1 for identification purposes, which implies that $\lambda_i = \lambda = \frac{1}{J}$.

For the prior variance, we combine data on familiarity and the γ parameter to form:

$$\delta_{ij0}^2 = \frac{1}{exp(\gamma_i Familiarity_{ij})}. \tag{D2}$$

The signal mean is given by brand intercepts, so will equal a vector of $\left[\mu_{i1}, \ldots, \mu_{iJ-1}, 1 - \sum_{j=1}^{J-1} \mu_{ij}\right]$, consistent with the restriction that brand intercepts sum to 1 (which is required for identification). In the Monte Carlo simulation, to simplify the estimation, we assume brand intercepts are constant, so will be given by a vector $[\mu, \ldots, \mu, 1 - (J-1) * \mu]$. In our estimation, brand intercepts are not constant and so the signal mean is given by the previous expression. The signal variance will be given by $\sigma_i^2 = \theta_i^2$.

Finally, search costs are obtained by combining parameters κ and ω with data on prior experience, as per:

$$c_{il} = exp(\kappa_i + \omega_i PriorExperience_{il}). \tag{D3}$$

With these variables in hand, we can form consumers' utilities from each search and choice decision they can make at every point in time. In particular, the set of available actions to the consumer at every point in time is given by $A = \{Search, Choice\}$, with an action $a \in Choice = \{1,\ldots,J\}$ denoting a brand, while an action $a \in Search = \{1,\ldots,J \times L\}$ denoting a brand-attribute combination. In our study there are 5 brands and 7 attributes, so the consumer can choose among 35 different possible brand-attribute combinations to search and 5 brands to purchase, resulting in a set A with 40 options available at every t.

The utility from stopping search and choosing an option j at time t is equal to (as per equation 6 in the paper): $v(\mu_{ijt}, \delta_{ijt}) + \epsilon_{ijt}$, where $(\mu_{ijt}, \delta_{ijt})$ are the updated mean and variance with information up to time t (using standard Bayesian updating rules - see equations 3 and 4 in the main paper), and ϵ_{ijt} is an idiosyncratic utility shock (T1EV distribution). The specific form of $v(\mu_{ijt}, \delta_{ijt})$ is given in the paper in equation 11. In contrast, the utility from searching brand j and attribute l equals (as per equation 7 in the paper): $-c_{il} + \max_{j,j' \in Choice}\{v(\mu_{ijt}, \delta_{ijt+1}) + \epsilon_{ijt}, v(\mu_{ij't}, \delta_{ij't}) + \epsilon_{ij't}\} + \eta_{ijt}$, where both $(\mu_{ijt}, \delta_{ijt})$ depend on the brand j and attribute l intended for search (and the search history) and where η_{ijt} is T1EV distributed. Using the expression for the expectation of the maximum of T1EV distributed variables, we obtain that the utility from searching brand j and attribute l is given by (as in equation 14)

$$-c_{il} + log\left[\sum_{j=1}^{J} exp(V(\mu_{ijt}, \delta_{ijt}))\right]. \tag{D4}$$

We additionally clarify the notation used in the expression above. For the brand intended for search j, the value inside the exponential equals $v(\mu_{ijt}, \delta_{ijt+1})$, i.e. the value of searching j, attribute l once more. For all other brands $j' \neq j$, the value inside the exponential equals $v(\mu_{ij't}, \delta_{ij't})$.

Our model assumes that a consumer picks the option that gives her the highest utility at every point in time. Utility maximization then allows us to write down the probability of observing a certain choice in our data, task to which we turn to now. More precisely, as shown in Section 5.4, a choice (either search or purchase) in our data occurs with probability (see equation 13 in the main paper–which we reproduce here)

$$P_{iat} = Prob(i \text{ takes action } a \text{ at } t) = \frac{exp(EV_{iat})}{\sum_{a' \in A} exp(EV_{ia't})}, \tag{D5}$$

where $EV_{iat} = v(\mu_{iat}, \delta_{iat})$ when $a \in Choice$ or EV_{iat} is equal to the value of searching a brand j and attribute l if $a \in Search$, i.e. it equals the expression in equation (D4) above.

Only one action will be taken at every time t. If this action is a search, then the consumers' prior mean and variance are updated (using Bayes' rule) and the whole process is repeated to generated choices at $t+1$. If this action is a choice, then the process stops as the consumer has ceased her search.

By repeating this process for all consumers, we generate our data set.

The entire discussion so far has focused on generating data with homogeneous parameters, i.e. parameters that do not vary across consumers. We next discuss how we can generate data where consumer parameters are heterogeneous. Towards this end, we specify N segments of consumers (in our case, $N = 2$), each with different parameters for preferences and search costs. Also, we specify a probability of each segment, such that the sum of the probabilities of each segment adds up to 1 (this probability approximates the size of each segment). For each segment, we then generate data on choices and prior beliefs as per the process described in the case of homogeneous parameters. Finally, we concatenate the choices made by each segment of consumers and save that as one data set. Appendix 13 will describe how we can then recover heterogeneous parameters from this data set.

13 Web Appendix E: Estimation routine

In this Appendix, we describe the estimation routine we use to recover preference and search cost parameters when data are generated as per the process described in Appendix 12. This routine will also be visible in the estimation code accompanying the paper.

Concretely, we aim to estimate the following set of parameters (same notation as in paper): μ which determines brand intercepts (signal mean), β which determines the effect of prior ownership, γ which determines the familiarity effect, θ which affects the signal variance, and κ and ω which determine the mean search cost and the effect of prior experience on search costs. To recover these parameters, we form a likelihood function describing consumer choices and try to maximize this likelihood in Matlab using an optimization tool called "fminunc". This optimization tool finds the minimum of unconstrained multivariable functions, so to maximize our likelihood function we will express it as a negative function that we then minimize using fminunc. The fminunc optimization tool works by successively attempting different sets of parameters, evaluating the likelihood function at those sets of parameters, and then determining, using a quasi-Newton algorithm, whether it is worth attempting another set of parameters.[34] Instead of computing the Hessian to determine whether an extreme point has been reached, the quasi-Newton method uses the function (the likelihood function in our case) and its derivative to approximate the Hessian, and then determines whether to try another set of parameters or stop because an extreme point was reached. We start the estimation with a set of parameters close to zero, but not equal to it to avoid boundary conditions: initial parameters will be approximately equal to those from the closest model to our own (the Weitzman model results in Table B-1), with values 0.1 for parameters that do not exist in the Weitzman model and for search costs. When estimating the model on simulated data, we start from a vector of 0.1 value for all parameters.

As mentioned, the estimation routine evaluates the likelihood function successively for different sets of parameters, given a certain starting value. We now describe how this likelihood function is constructed given data on consumer choices (using the same structure as in Figure D-1) and a set of parameters to evaluate. For every consumer i, retrieve data on their choices (searches in every

[34] More details on Matlab's fminunc optimization tool can be found at `https://www.mathworks.com/help/optim/ug/fminunc.html`.

time period and the final brand choice) and prior information (prior ownership, familiarity, prior experience). Combine these data and the set of parameters to form the variables of interest that affect consumer choices. This process will follow the same steps as in Appendix 12 above, as well as in our main paper. More precisely, for the prior mean, we combine data on prior ownership and the β parameter to form:

$$\mu_{ij0} = \lambda_i + \beta_i PriorOwnership_{ij}, \tag{E1}$$

for every consumer-brand combination. As mentioned in the paper, λ_i is given by the average of the brand intercepts, so $\lambda_i = \frac{\sum_{j=1}^{J} \mu_{ij}}{J}$. In Section 5.5, we impose the restriction that brand intercepts sum to 1 for identification purposes, which implies that $\lambda_i = \lambda = \frac{1}{J}$.

For the prior variance, we combine data on familiarity and the γ parameter to form:

$$\delta_{ij0}^2 = \frac{1}{exp(\gamma_i Familiarity_{ij})}. \tag{E2}$$

The signal mean is given by brand intercepts, so will equal a vector of $\left[\mu_{i1}, \ldots, \mu_{iJ-1}, 1 - \sum_{j=1}^{J-1} \mu_{ij}\right]$, consistent with the restriction that brand intercepts sum to 1 (which is required for identification). The signal variance will be given by $\sigma_i^2 = \theta_i^2$.

Finally, search costs are obtained by combining parameters κ and ω with data on prior experience as per:

$$c_{il} = exp(\kappa_i + \omega_i PriorExperience_{il}). \tag{E3}$$

For every consumer i and every choice occasion t, we determine the index of the choice she made. In other words, we determine whether she searched a brand-attribute combination or whether she stopped her search and chose a brand. Our model assumes that consumers make the choice that maximizes their utility, net of search costs, at every point in time. Thus, if the consumer chose to search an option (brand-attribute combination) at t, we know that her utility from searching this option must be higher than the utility from searching other options or stopping. Similarly, if the consumer stops searching, then we know that the utility from the chosen option is greater than the utility of continuing her search. To compute the probability of a certain action and thus the likelihood function, all we need is to write down consumers' utility from searching and stopping. The utility from stopping search and choosing an option j at time t is equal to (as per equation 6 in the paper): $v(\mu_{ijt}, \delta_{ijt}) + \epsilon_{ijt}$, where $(\mu_{ijt}, \delta_{ijt})$ are the updated mean and variance with information up to time t (using standard Bayesian updating rules - see equations 3 and 4 in the main paper), and ϵ_{ijt} is an idiosyncratic utility shock (T1EV distribution). The specific form of $v(\mu_{ijt}, \delta_{ijt})$ is given in the paper in equation 11. In contrast, the utility from searching brand j and attribute l equals (as per equation 7 in the paper): $-c_{il} + \max_{j,j' \in Choice}\{v(\mu_{ijt}, \delta_{ijt+1}) + \epsilon_{ijt}, v(\mu_{ij't}, \delta_{ij't}) + \epsilon_{ij't}\} + \eta_{ijt}$, where both $(\mu_{ijt}, \delta_{ijt})$ depend on the brand j and attribute l intended for search (and the search history) and where η_{ijt} is T1EV distributed. Using the expression for the expectation of the maximum of T1EV distributed variables, we obtain that the utility from searching brand j and attribute l is given by (as in equation 14)

$$-c_{il} + log\left[\sum_{j=1}^{J} exp(V(\mu_{ijt}, \delta_{ijt}))\right]. \tag{E4}$$

We additionally clarify the notation used in the expression above. For the brand intended for search j, the value inside the exponential equals $v(\mu_{ijt}, \delta_{ijt+1})$, i.e. the value of searching j, attribute l once more. For all other brands $j' \neq j$, the value inside the exponential equals $v(\mu_{ij't}, \delta_{ij't})$.

Having assumed T1EV utility error shows, a choice (either search or purchase) in our data occurs with probability (see equation 13 in the main paper–which we reproduce here)

$$P_{iat} = Prob(i \text{ takes action } a \text{ at } t) = \frac{exp(EV_{iat})}{\sum\limits_{a' \in A} exp(EV_{ia't})}, \tag{E5}$$

where

$$EV_{iat} = \begin{cases} v(\mu_{iat}, \delta_{iat}) & \text{if } a \in Choice \\ -c_{il} + log\left[\sum_{j=1}^{J} exp(v(\mu_{ijt}, \delta_{ijt}))\right] & \text{if } a \in Search \ \& \ a = (j, l). \end{cases} \tag{E6}$$

where $a \in Choice = \{1, \ldots, J\}$ denotes a brand chosen, while an action $a \in Search = \{1, \ldots, J \times L\}$ denotes and brand-attribute combination searched.

After each search action taken, the consumer observes a signal from the distribution $N(\mu_{ij}, \sigma_i^2)$. Therefore, in addition to consumer preference and search cost parameters, the specific history of the signals obtained up to time t can affect choices at t. These signals are unobserved by the researcher, but will affect P_{iat}. Therefore, to integrate over the distribution of signals consumers obtain while searching, we draw S possible signal histories for each consumer, brand, attribute, and time of search, and average the resulting choice probabilities to form P_{iat}. Because these choice probabilities do not have a closed-form solution after averaging over signal draws, we will use the simulated log-likelihood function to estimate parameters instead. This function is given by

$$SLL = \sum_{i} \sum_{a} \sum_{t} d_{iat} log(P_{iat}), \tag{E7}$$

where $d_{iat} = 1$ if consumer i chooses action a at t, and zero otherwise.

As mentioned above, the negative of the value SLL will be used in the fminunc optimization routine in Matlab to recover parameters.

The entire discussion so far has focused on recovering homogeneous parameters, i.e. parameters that do not vary across consumers. We next discuss how we can recover consumer heterogeneity in parameters using a latent class approach. This approach hypothesizes the existence of N segments of consumers in the data, where N is unknown to the researcher (hence, latent). Consistent with our main paper, in this section we will describe the estimation procedure for $N = 2$, although these same steps can be applied for any N.

Suppose $N = 2$ and there are k parameters determining choices for each segment. Also, suppose $\pi_1 = f(\rho)$ gives the probability of belonging to segment 1, which is a function of the parameter ρ that needs to be estimated (in our empirical specification, we use $f(\rho) = exp(\rho)/(1 + exp(\rho))$ to ensure that this probability is positive and falls within the interval [0,1]). The probability of belonging to segment 2 is then given by $1-\pi_1$. This results in a total of $N \times k+1$ parameters to estimate, including the probability of belonging to a segment. For the set of parameters $n \in \{1,2\}$, we will form variables of interest as above (prior mean, prior variance, signal mean, signal variance, and search costs) and compute P^n_{iat} for n before averaging over the signal draws. We then compute a weighted value of the P^n_{iat}'s using the probability of a consumer belonging to segment n and obtain $P_{iat} = P^1_{iat} \times \pi_1 + P^2_{iat} \times (1-\pi_1)$. With P_{iat} in hand, we then follow the same steps as above to integrate over signal draws and form the simulated likelihood. By maximizing this simulated likelihood function, we are able to find the set of $N \times k+1$ parameters that describe behavior in the case of consumer heteorogeneity.

14 Web Appendix F: Additional identification details

In this section, we illustrate how the selection rule allows us to separately identify the mean search cost parameter from the signal variance. We will do so using an example. After discussing this example, which is based on our current model, we also sketch identification arguments for two possible extensions of our model.

Suppose there is one consumer, two brands, A and B, one attribute, and no outside option. Also suppose that the consumer searches A first and then searches B, each for one fixation. Let the search cost equal c.[35] For simplicity, we write the prior mean as μ_{j0} and the prior variance as δ^2_{j0}, with the understanding that these values vary across brands as a function of prior information (prior ownership and familiarity). Finally, we abstract away from the role that the unobserved error terms play in this example, since their presence will not affect our identification argument.

At $t=0$, from the selection rule, the decision to search product A first implies the following restriction on our parameters of interest:[36]

$$-c+\max\{v(\mu_{A0},\delta_{A1}), v(\mu_{B0},\delta_{B0})\} > -c+\max\{v(\mu_{A0},\delta_{A0}), v(\mu_{B0},\delta_{B1})\}, \tag{F1}$$

where the left hand side gives the value of searching brand A, while the right hand side gives the value of searching B at $t=0$. We can immediately see that the search cost drops out of the selection rule. The same will be true of the mean search cost in our empirical model.

From the Bayesian updating rules in equation (4) we know that

$$\delta^2_{j1} = \frac{\delta^2_{j0}\sigma^2}{\delta^2_{j0}+\sigma^2}, \tag{F2}$$

where σ^2 denotes the signal variance. Since $\delta_{j1} < \delta_{j0}$ and utility is decreasing in uncertainty, it follows that $v(\mu_{A0},\delta_{A1}) > v(\mu_{A0},\delta_{A0})$ and $v(\mu_{B0},\delta_{B1}) > v(\mu_{B0},\delta_{B0})$. In addition, inequality (F1) reveals other

[35] We assume, for simplicity, that there is only one attribute, making the prior experience redundant. We made these assumptions in order to focus on the main challenge which involves the separate identification of mean search costs and the mean signal variance.

[36] With slight abuse of notation, we will write all inequalities as strict inequalities, which will allow us to simplify the discussion without having to assume an arbitrary direction in which inequalities are not strict.

relations of interest. First, it reveals that $v(\mu_{A0}, \delta_{A1}) > v(\mu_{B0}, \delta_{B0})$ since otherwise it would need to be the case that $v(\mu_{B0}, \delta_{B0}) < v(\mu_{B0}, \delta_{B1})$, which is a contradiction. Then, if $v(\mu_{A0}, \delta_{A1}) > v(\mu_{B0}, \delta_{B0})$, there are two cases to consider. Either $v(\mu_{A0}, \delta_{A0}) > v(\mu_{B0}, \delta_{B1})$ or the opposite. In both cases it follows that

$$v(\mu_{A0}, \delta_{A1}) > v(\mu_{B0}, \delta_{B1}). \tag{F3}$$

At $t = 1$, after searching A, the consumer observes a signal s_1 on A drawn from the distribution $N(\mu_A, \sigma^2)$. With this signal, the consumer updates her posterior mean using Bayesian updating rules in equation (3). More precisely, this equals

$$\mu_{A1} = \frac{\frac{\mu_{A0}}{\delta_{A0}^2} + \frac{s_1}{\sigma^2}}{\frac{1}{\delta_{A0}^2} + \frac{1}{\sigma^2}}, \tag{F4}$$

while the posterior variance is updated to δ_{A1}^2. The mean and variance for brand B remain unchanged since this brand was not searched (they equal (μ_{B0}, δ_{B0})).

If the consumer prefers to switch and search B at $t = 1$, then from the selection rule we know

$$-c + \max\{v(\mu_{A1}, \delta_{A2}), v(\mu_{B0}, \delta_{B0})\} < -c + \max\{v(\mu_{A1}, \delta_{A1}), v(\mu_{B0}, \delta_{B1})\}. \tag{F5}$$

Since $\delta_{jt+1} < \delta_{jt}$ and utility is decreasing in uncertainty, it follows that $v(\mu_{A1}, \delta_{A2}) > v(\mu_{A1}, \delta_{A1})$ and $v(\mu_{B0}, \delta_{B1}) > v(\mu_{B0}, \delta_{B0})$.

Searching B at $t = 1$ reveals that

$$v(\mu_{A1}, \delta_{A1}) < v(\mu_{B0}, \delta_{B1}). \tag{F6}$$

To see this, suppose the opposite was true: $v(\mu_{A1}, \delta_{A1}) > v(\mu_{B0}, \delta_{B1})$. There are two cases to consider: either $v(\mu_{A1}, \delta_{A2}) > v(\mu_{B0}, \delta_{B0})$ or the opposite. In the first case, from (F5) it follows that $v(\mu_{A1}, \delta_{A2}) < v(\mu_{A1}, \delta_{A1})$, which is a contradiction. In the second case, if $v(\mu_{A1}, \delta_{A1}) > v(\mu_{B0}, \delta_{B1})$ and $v(\mu_{A1}, \delta_{A2}) < v(\mu_{B0}, \delta_{B0})$, then from (F5) it must be $v(\mu_{B0}, \delta_{B0}) < v(\mu_{A1}, \delta_{A1})$, which implies $v(\mu_{A1}, \delta_{A2}) < v(\mu_{A1}, \delta_{A1})$, which is a contradiction. Therefore, we can conclude that inequality (F6) holds.

Inequalities (F3) and (F6) reveal how the signal variance is identified from switches in the search patterns observed.[37] To see this, notice the fact that the right hand side of both inequalities is the same, i.e. they equal $v(\mu_{B0}, \delta_{B1})$.[38] The left hand side is a function of the value of searching A. If μ_{A1} was similar to or larger than μ_{A0}, then the inequality (F6) could not hold, since expected utility is increasing in the posterior mean. Therefore, the fact that the consumer preferred to switch and not search A again implies that the signal obtained after searching A lowered her posterior mean, making μ_{A1} lower than μ_{A0}. The inherent randomness in the signal draw, the sequence of signals observed, as well as consumers' search switching behavior will identify the effect of σ separately from the effect of the mean search cost which drops out of the selection rule (as can be seen above).

The rest of this appendix will sketch identification arguments for two extensions of our model.

[37]Other inequalities we derived above may also contribute to this identification.

[38]This right hand side value is the threshold we refer to in the main text, in Section 5.5.

First, we consider the case where search costs are brand-specific. This case could arise in our empirical setting when observables such as the location of the product on the screen affected search costs (as in Ursu (2018)). For example, in this case we could write $c_j = exp(\kappa + \gamma * Loc_j)$. The search cost mean κ would be separately identified from the signal using the same argument as above. The brand-specific search costs γ would be identified from variation in the data, i.e. in the location of the product on the screen, and the frequency with which such products are searched. For example, if brands located at the top of the screen are searched more frequently, then we would infer they have lower search costs. For similar arguments, see Ursu (2018). Any other observable brand-specific variables that affected search costs would be identified in the same way.

The second extension we will consider involves the case where search costs vary over the search process with t. This case could arise if search costs were a function of the distance the eye traveled from one eye-fixation to the next. For example, it could be the case that moving ones eyes to a further location was more costly than looking at the same location as on the previous fixation. Such search costs arise in other settings, for example when transportation costs affect the search process (see Yavorsky et al. (2021)). If distance was exogenous (as in our case where brand and attribute information were randomized across conditions), then this variation would allow us to separately identify the t-specific component of the search cost from other parameters using the selection and the stopping rules. Also, the mean search cost would be separately identified from the signal variation using the same argument as above.

We have attempted to estimate versions of our model with search costs varying across brands and across t. Unfortunately, the complexity of our model did not allow us to obtain reliable estimates in these cases.

15 Web Appendix G: Numerical issues

As mentioned in Section 5.4, numerical issues can arise in our empirical setting when estimating the model. The primary reason for such issues is the large number of observed searches in our data. More precisely, in our estimation sample the average (median) number of searches equals 142 (131). In contrast, in prior work that estimates a search model, the number of searches rarely exceeds 3. Thus, in our setting, summing $log(P_{iat})$ over t for consumers for whom t is large (for example larger than 200) may result (for some parameter value attempts) in a value for the likelihood function that approaches negative infinity. This issue is compounded by the fact that consumers in our setting also choose among a large set of options (40 in total), which means the probability P_{iat} is generally small, making the $log(P_{iat})$ a relatively larger negative number, which, when summed up over a large t, is even more likely to approach negative infinity. Since the likelihood function does not vary with changes in the parameter values when it equals negative infinity, consumers with large values of t will not contribute to the likelihood calculation.

To address these numerical issues in our estimation, we have followed the recommendations outlined in Conlon and Gortmaker (2020) for best practices when facing numerical estimation problems, as well as have tried a number of other approaches. In particular, among the methods we tried were the following:

1. Rewrite the likelihood function using the log-sum-exp trick which prevents it from approaching infinity: $LSE(x) = log\sum_k exp(x_k) = a + log\sum_k exp(x_k - a)$ where $a = max\{0, \max_k x_k\}$.

2. Impose box constraints on the parameter space to prevent the model from attempting parameter values that will lead to the issues discussed above.

3. Tried multiple starting values (such as starting the estimation at zeros, rather than the Weitzman parameter results).

4. Tried estimating the model only on data from consumers who made fewer than 100 searches.

5. Fixed certain parameters and only estimated the remaining parameters.

6. Increased the number of signal draws to 100 (from 50).

Across all methods, our main qualitative findings (the fact that our main model fits the data better and that brand intercepts are inflated when prior information is not accounted for) continue to hold. The only exception involves the first method above that generally recovers similar parameters for the model without prior information, but that results in search costs that tend to zero (on the order of $exp(-25)$), large standard errors (exceeding 100), and inconsistent results for the Samsung and Apple brand intercepts in our main model specification. What continues to be true is that our main model fits the data better than the model that ignores prior information. Given these findings, we believe that the results we presented in our main paper are the most robust, given the numerical issues we have described in this section.

16 Web Appendix H: Supporting the one-step ahead decision rule

As mentioned in Section 5, our proposed decision rule lies in between a myopic and an optimal fully forward-looking rule. In this section, we provide more explanation for our claim.

In our model, the consumer searches as if she has to make a purchase decision right after the immediate search, i.e. as if the next search decision is her last. That is, the consumer uses a one-step ahead decision rule: she decides whether to stop or search once more and then stop her search. In this model, if the consumer decides to continue searching, she takes into account the consequences of one more search on her future decisions.

In contrast, a myopic rule is one under which the consumer does not take into account how her current decision may affect the evolution of her future beliefs and thus any future searches or purchase decisions. Rather, under a myopic rule, the consumer makes decisions at some time t based on the highest per-period utility obtained so far (with information up to t). This model is referred to as the Bayesian-myopic rule and has been frequently used in prior Marketing and Economics work (see Coscelli and Shum (2004), Narayanan et al. (2005), Chintagunta et al. (2009), Dickstein et al. (2018)). Mathematically, in our model this decision rule would reduce to

- At every t, choose

$$\max_{j \in Choice} v(\mu_{ijt}, \delta_{ijt}) + \epsilon_{ijt}. \tag{H1}$$

This model cannot be used to explain both search and purchase decisions in our setting. Rather, our setting requires a model that can capture the tradeoff consumers make between stopping search now and continuing search, taking into account the consequences of such actions on future decisions. The model we propose is one where a consumer who decides to continue searching takes into account how her decision will affect her future utility (over the next period) and therefore her future decisions.

A fully forward-looking rule would model consumers as making decisions taking into account how these actions affect all their future decisions, not only their next period decisions. We can illustrate the added complexity of such a model by looking at the relatively simpler case where the consumer considers the consequences of her actions over two future periods, rather than all future periods. Now, at some point t, the consumer can choose between three options: stop search, search once more and then stop searching, or search two more times and then stop searching. Mathematically, we can write:

At t, the consumer faces three options:

1. Stop searching and **choose**

$$\max_{j \in Choice} v(\mu_{ijt}, \delta_{ijt}) + \epsilon_{ijt}. \tag{H2}$$

2. **Search** product j and attribute l and then stop to receive the maximum utility derived from choosing one of the alternatives immediately thereafter

$$\max_{(j,l) \in Search} \{-c_{il} + \max_{j,j' \in Choice} \{v(\mu_{ijt}, \delta_{ijt+1}) + \epsilon_{ijt}, v(\mu_{ij't}, \delta_{ij't}) + \epsilon_{ij't}\} + \eta_{ijt}\}. \tag{H3}$$

3. **Search** product k and attribute m at t, search product j and attribute l at $t+1$, and then stop to receive the maximum utility derived from choosing one of the alternatives immediately thereafter

$$\max_{(k,m) \in Search} \{-c_{im} + \int_{-\infty}^{\infty} [\max\{\max_{k \in Choice} v(\mu_{ikt+1}, \delta_{ikt+1}) + \epsilon_{ikt+1},$$
$$\max_{(j,l) \in Search} \{-c_{il} + \max_{j,j' \in Choice} \{v(\mu_{ijt+1}, \delta_{ijt+2}) + \epsilon_{ijt+1}, v(\mu_{ij't+1}, \delta_{ij't+1}) + \epsilon_{ij't+1}\} + \eta_{ijt+1}\}\}] g(s) d(s). \tag{H4}$$

where the integral in the third option accounts for the signal s the consumer observes when searching j and attribute l at t. The first two options mirror those in our model. The consumer's third option of performing two more searches can be understood as follows. The consumer can search k and attribute m at t. Then, she will observe a signal s, with which she will update her utility to $v(\mu_{ikt+1}, \delta_{ikt+1}) + \epsilon_{ikt+1}$. Given that this signal is unobserved by the consumer before searching (also unobserved by the researcher), we need to integrate over all possible signals when estimating the model. Then, at $t+1$, she can either stop searching and get the highest observed utility so far (note that $\max_{k \in Choice}$ contains all products available for purchase, not only k). Or, she can continue searching, choosing product j and attribute l to search next (including the option to search k and attribute m again). If she does so, then the consumer needs to take into account how her second search will affect her future utility.

Continuing to add additional steps to the problem described above will bring the model closer to the fully forward-looking model. A fully forward-looking model of search would take into account all possible future periods and derive an optimal search rule for a rational consumer. Unfortunately, the added complexity of having to integrate over unobserved signals over multiple periods, makes it difficult for us to determine the optimal rule in this model. Therefore, we rely on prior work that has obtained experimental evidence for the superiority of the one-step ahead rule to support our proposed one-step ahead model (see Camerer and Johnson (2004), Gabaix et al. (2006), Tehrani and Ching (2020)). We are also encouraged by the fact that prior empirical work in Marketing and Economics has successfully applied one-step ahead models in settings where consumers search (see Hodgson and Lewis (2020), Ursu et al. (2022)). Finally, we would like to recall that in our empirical setting, consumers search by simply moving their eyes and making another fixation every 200-400 milliseconds. In particular, given the high frequency with which consumers make search decisions in our setting (the average consumer makes 229 fixations), we expect the computational burden of determining the optimal eye fixation sequence for all future periods to outweigh the benefit, leading consumers to use heuristics over optimal search rules. For all these reasons, we believe our one-step ahead decision rule to be a good proxy for consumer choices in our setting. Also, for computational reasons, we are unable to extend the estimation of our model even to the two-step ahead rule, since this would require solving the complicated integral in the problem presented above. As mentioned in Section 8, deriving search rules for our problem assuming consumers optimally look $n > 1$ steps into the future can be an area of future research.