WHAT DO CONSUMERS CONSIDER BEFORE THEY CHOOSE? IDENTIFICATION FROM ASYMMETRIC DEMAND RESPONSES*

Jason Abaluck and Abi Adams-Prassl

Consideration set models generalize discrete-choice models by relaxing the assumption that consumers consider all available options. Determining which options were considered has previously required either survey data or restrictions on how attributes affect consideration or utility. We provide an alternative route. In full-consideration models, choice probabilities satisfy a symmetry property analogous to Slutsky symmetry in continuous-choice models. This symmetry breaks down in consideration set models when changes in characteristics perturb consideration. We show that consideration probabilities are constructively identified from the resulting asymmetries. We validate our approach in a lab experiment where consideration sets are known and then apply our framework to study a "smart default" policy in Medicare Part D, wherein consumers are automatically reassigned to lower-cost prescription drug plans with the option of opting out. Full-consideration models imply that such a policy will be ineffective because consumers will opt out to avoid switching costs. Allowing for inattention, we find that defaulting all consumers to lower-cost options produces negligible welfare benefits on average, but defaulting only consumers who would save at least \$300 produces large benefits. *JEL Codes:* D12, D90, I11.

I. Introduction

Discrete-choice models generally assume that consumers consider all available options when making their choices. This

*Thanks to Leila Bengali, Mauricio Caceres Bravo, and Anna Papp for excellent research assistance and to Dan Ackerberg, Joe Altonji, Dan Benjamin, Doug Bernheim, Steve Berry, Richard Blundell, Judy Chevalier, Ian Crawford, Francis Di Traglia, Aureo de Paula, Jonathan Feinstein, Jeremy Fox, Xavier Gabaix, Jonathan Gruber, Phil Haile, Hamish Low, Erzo Luttmer, Paola Manzini, Marco Mariotti, Olivia Mitchell, Francesca Molinari, Fiona Scott Morton, Barry Nalebuff, Jeremias Prassl, Alan Sorensen, Joe Shapiro, K. Sudhir, Chris Taber, and participants in the Heterogeneity in Supply and Demand Conference, the Roybal Annual Meeting, and seminar participants at Berkeley, Harvard, LSE, Oxford, Stanford, the University of Chicago, Wharton, and Yale. We acknowledge financial support from National Institute on Ageing grant no. R01 AG031270 and the Economic and Social Research Council, Grant ES/N017099/1. We would like to thank the editor and four anonymous referees for their help in improving the article. Any remaining errors are rationalizable with sufficiently flexible preferences.

The Quarterly Journal of Economics (2021), 1611–1663. doi:10.1093/qje/qjab008.
Advance Access publication on March 13, 2021.

prevents researchers from asking many questions of interest. What factors lead consumers to become aware of more options? Will inertial consumers "wake up" in response to a price increase but remain unresponsive if rivals lower prices? Normatively, whether people choose the same products year after year because they like those options or because they do not know what else exists has first-order consequences for welfare. If we can measure preferences conditional on consideration, we can assess the benefits of policies that make consumers aware of more options or redirect consumers to products they would choose if they considered them. Such policies are ubiquitous, ranging from defaulting people into lower-cost insurance plans to populating online shopping carts with items that people might like.

Consideration set models are a generalization of discrete-choice models that relax the assumption that individuals consider all goods. These models instead specify a probability that each subset of options is considered (Manski 1977). The approach has long been applied in the marketing literature (Hauser and Wernerfelt 1990; Shocker et al. 1991) and has become increasingly popular in both theoretical and applied literatures in economics. Consideration sets might arise due to inattention or bounded rationality (Treisman and Gelade 1980), from search costs (Caplin, Dean, and Leahy 2019), or because consumers face (unobserved) constraints on what options can be chosen (Gaynor, Propper, and Seiler 2016).[1] In contrast to tests of rationality, such as checking whether consumers make dominated choices, consideration set models allow us to simulate how consumers would choose if they were informed about relevant options.

Identification is an immediate concern in consideration set models—if changes in prices or other characteristics perturb demand, can we tell whether this effect comes via consideration or utility? The results in this article highlight a new source of identifying variation in two widely used classes of consideration set models that have been the focus of much applied and theoretical work. In the first class of model, which we call the default

1. Given the two-stage framework in this article, "attention," "awareness," and "consideration" all synonymously mean "a good is in the choice set from which consumers then maximize utility." We assume that conditional on considering a good, one observes all of its relevant attributes. For theoretical frameworks that relax this assumption see, for example, Kőszegi and Szeidl (2012), Bordalo, Gennaioli, and Shleifer (2013), and Gabaix (2014), among others.

specific consideration (DSC) model, consumers are either “asleep” and choose a default option or they “wake up” and make an active choice from all products (Ho, Hogan, and Scott Morton 2017; Hortaçsu, Madanizadeh, and Puller 2017; Heiss et al. 2016). In the second class of model, which we call the alternative specific consideration (ASC) model, each good has an independent consideration probability that depends on characteristics of the good in question (Swait and Ben-Akiva 1987; Ben-Akiva and Boccara 1995; Goeree 2008; Van Nierop et al. 2010; Manzini and Mariotti 2014; Kawaguchi, Uetake, and Watanabe forthcoming).

Empirical models of both types usually rely either on auxiliary data on what goods are considered or on additional exclusion restrictions for point identification of the structural functions of interest. These exclusion restrictions are often questionable and can be in tension with economic theory; excluding prices from consideration, for example, can be inconsistent with simple models of rational inattention, and it is unclear that advertising only affects choices via informing consumers about which goods exist (Goeree 2008; Van Nierop et al. 2010). Yet agnosticism over what variables affect utility and which affect attention is usually associated with only partial identification of the objects of interest (Lu 2016; Barseghyan et al. 2021).

Our approach builds on recent literature in behavioral decision theory on limited consideration models (Masatlioglu, Nakajima, and Ozbay 2012; Manzini and Mariotti 2014; Cattaneo et al. 2020). We prove that the restrictions on choice probabilities that are already imposed in most settings are sufficient for point identification of both preferences and consideration probabilities in the DSC and ASC frameworks, as well as hybrid models combining features of both alternatives. Our method does not require auxiliary information on consideration sets, and it allows all observables to affect consideration and utility. We provide simple closed-form expressions for consideration set probabilities in terms of differences in cross-derivatives (the discrete-choice analog of Slutsky asymmetries). Our framework subsumes many of the consideration set models in the applied literature and does not rely on assuming a particular functional form for random utility errors. In cross-sectional data, our results can be used to identify whether goods are demanded because they are high utility or because they are more likely to be considered. In panel data, one can evaluate whether inertia reflects utility-relevant factors or inattention. More generally, in the class of models we describe, one can ask how consumers would choose with full consideration, and one can do so with no additional data beyond what is required to estimate conventional discrete-choice models.

Our identification result builds on the insight that imperfect consideration breaks symmetry between cross-price responses (or more generally, cross-characteristic responses). For example, in a model with a default, symmetry would ordinarily require that switching decisions be equally responsive to an increase in the price of the default good by $100 or a decrease in the price of all rival goods by $100. Suppose instead that consumers are inattentive and choose the default option unless that good becomes sufficiently unsuitable. Now switching decisions will be unresponsive to changes in the price of rival goods but more responsive to changes in the price of the default to the degree that these changes perturb attention (Moshkin and Shachar 2002). Although the link between imperfect attention and Slutsky asymmetry has been discussed in the theoretical literature, notably in Gabaix (2014), and noted as a source of identifying variation in Moshkin and Shachar (2002), this approach has not yet been developed in the generality we consider.[2] Our framework implies that attempts to model consideration sets such as fixed effects in utility for products on different shelves or interactions between prices and fixed effects can still yield misspecified models because they do not relax the symmetry assumption.

Our identification proof is constructive and so, in theory, consistent nonparametric estimators could be directly based on it. However, in most applications of interest, we advocate estimating parametric generalizations of conventional models.[3] We consider two estimation approaches: indirect inference and maximum likelihood. Maximum likelihood estimation makes the link between estimation and identification less explicit but can be computationally more feasible. To estimate the model by indirect inference, we specify a flexible auxiliary model that permits a general pattern of asymmetries and then estimate the parameters of our consideration set model to fit them.

2. Aguiar and Serrano (2017) use deviations from Slutsky symmetry to quantify violations of rationality but do not use these for constructive identification of behavioral phenomena.

3. A STATA command that implements several special cases of our model is available for download as “alogit”; a user’s guide and sample data sets can be downloaded at https://sites.google.com/view/alogit/home.

We validate our approach in a lab experiment in which participants made a series of choices from (known) proper subsets of 10 possible goods. Using only data on choices and ignoring information on what items were considered, matching Slutsky asymmetries enables us to accurately recover the probabilities that each good was available as well as recovering the preference parameters that we would estimate conditional on knowing which items were available. Conventional models with a comparable number of parameters misspecify own- and cross-price elasticities relative to the elasticities computed using data on which items were actually available. We formally test whether our consideration set model can generate the asymmetries captured by our flexible auxiliary model, finding that our framework cannot be rejected while ad hoc generalizations of full-information models (e.g., allowing for good-specific price effects in a standard conditional logit) cannot explain the reduced-form patterns.

We apply our framework to analyze prescription drug insurance choices in Medicare Part D. We allow for consumers to be completely "asleep" (and simply choose their default plan) and for consumers to attend to only a subset of options even in periods where they actively search. In this market, more than 90% of beneficiaries are inertial. This has led to proposals for a "smart default" policy, in which consumers are automatically switched into lower-cost plans but can opt out (Handel and Kolstad 2015). Evaluating this policy from both positive and normative perspectives requires us to disentangle the degree to which inertia in plan choice is driven by limited consideration or utility-relevant switching costs. If beneficiaries are inertial because it is costly for them to acclimate to a new plan, then most beneficiaries will opt out of the smart default. In fact, models with full attention imply that defaults have no effect at all on choices,[4] and models with some inattention but high acclimation costs imply that inattentive consumers are affected but made worse off by having to pay these costs. Alternatively, if beneficiaries are inertial largely due to limited consideration and have low acclimation costs, they may

4. In Online Appendix C, we examine the robustness of our results to allowing for "paperwork costs," that is, costs to choose a plan different from the default regardless of whether you have enrolled in that plan already. With paperwork costs, defaults may be sticky in full-consideration models. We find that our welfare conclusions are robust to the existence of paperwork costs as these costs are small relative to acclimation costs.

be made better off by being defaulted to a lower-cost plan. When we allow for limited consideration, we find a high degree of inattention and positive welfare effects of smart defaults, especially if we only switch consumers with cost savings at least as great as estimated acclimation costs.

The rest of this article proceeds as follows. Section II situates our approach in the existing literature. Section III lays out our general model and identification results. Section IV validates our approach in the lab where we observe consideration and develops an indirect inference estimator in which structural parameters are chosen to match cross-derivative asymmetries in the data. Section V harnesses our framework to estimate consumer preferences and limited consideration in Medicare Part D and conducts a welfare analysis of a smart default policy in this setting. Section VI concludes.

II. Related Literature

1. Exclusion Restrictions. Identification of consideration set models is typically achieved by restricting which variables can influence consideration and utility. In general, point identification of all structural functions requires one to exclude a set of variables that affect consideration from utility and a set of variables that affect utility from consideration.[5] Our identification result does not rely on the existence of variables that influence consideration but not utility and vice versa. This provides a route for researchers to test generally which factors are important for consideration and utility, enabling one to, for example, distinguish between models of naive versus rational consideration. This contrasts with empirical strategies that exclude prices from consideration, and thus cannot allow consideration to be driven by the expected benefits of search (Kawaguchi, Uetake, and Watanabe forthcoming; Goeree 2008).

2. Auxiliary Data. A second strand of literature identifies consideration set models using auxiliary data on which goods were considered. Conlon and Mortimer (2013) assume that

5. For example, Goeree (2008), Hortaçsu, Madanizadeh, and Puller (2017), Gaynor, Propper, and Seiler (2016), and Heiss et al. (2016) proceed in this way. For formal results, see the literature on the identification of mixture models (Compiani and Kitamura 2016). Barseghyan, Molinari, and Thirkettle (2021) show that one only requires variables that influence utility to be excluded from consideration in combination with an "identification at infinity" type argument for identification of an ASC type model.

consideration sets are known in some periods, Draganska and Klapper (2011) and Honka and Chintagunta (2016) use survey data on what products are and are not considered when choosing, and Reutskaja et al. (2011) use eye-tracking methods to follow what options individuals consider. However, there are many scenarios where such auxiliary data do not exist but limited consideration is a first-order concern. Furthermore, many process-tracking procedures measure attentional inputs but not consideration itself. For example, we may observe the rank of products in search or perhaps eye-tracking software. As noted by Gabaix (2019), it is important to treat these as correlates of consideration rather than a direct measure of attention itself, that is, to add such variables as determinants of the (unobserved) consideration probability. Our framework enables researchers to do this.

3. Theoretical Restrictions. We are able to relax the assumptions usually required for identification of consideration set models by exploiting the restrictions implied by standard assumptions made in discrete-choice analysis. There is a growing body of literature in behavioral decision theory that highlights the identifying power of theoretical restrictions in consideration set models (Masatlioglu, Nakajima, and Ozbay 2012; Manzini and Mariotti 2014; Cattaneo et al. 2020). These papers show that changes in choice probabilities that result from changes in the set of available products (an exogenous potential change to a consumer's consideration set) place restrictions on the set of preferences and consideration sets that can rationalize the data. Manzini and Mariotti (2014) prove that consideration probabilities and a consumer's preference relation can be uniquely identified from individual choice data if one observes choice from every possible nondegenerate subset of feasible alternatives, while model primitives in Masatlioglu, Nakajima, and Ozbay (2012) and Cattaneo et al. (2020) remain only partially identified in general. These insights have only been directly harnessed in experimental work in which it is possible to generate such variation (Aguiar et al. 2018).

Kawaguchi, Uetake, and Watanabe (forthcoming) make the weaker assumption of "leave-one-out" variation in product availability to identify a consideration set model to study optimal product recommendations. Their condition relates the percentage change in product demand when a single product is unavailable to consideration probabilities using an identification at infinity argument made possible by excluding price from consideration. Our results for the ASC and hybrid models rely on variation equivalent to Kawaguchi, Uetake, and Watanabe (forthcoming) but without the need for additional exclusion restrictions.

Our identification result harnesses the identifying power of deviations from Slutsky symmetry.[6] In a version of the DSC model that we consider herein, Moshkin and Shachar (2002) show that switching probabilities are more sensitive to changes in the characteristics of default plans than to nondefault plans. In this article, we prove that this variation is sufficient for identification of consideration probabilities given the assumptions made in many discrete settings and show that similar insights extend to a much richer class of models than that focused on by Moshkin and Shachar (2002).

4. Partial Identification. Although our results do not rely on additional exclusion restrictions, we do work within the structure imposed by popular models of consideration set formation. Permitting preference heterogeneity over alternatives in the population, Barseghyan et al. (2021) leave the process generating consideration sets completely unrestricted and allow for dependence between unobservables driving consideration and utility. This approach accommodates a wider set of limited-attention models than our approach, at the cost of losing point identification of the structural functions of interest. Our contribution is to show that in a framework that encompasses many specifications currently estimated in the applied economics and marketing literature, all structural functions of interest are point identified from choice probabilities and cross-price elasticities.

5. Alternative Assumptions on Preferences. A final strand of the literature restricts the nature of preferences and preference heterogeneity for the purposes of identification. Crawford, Griffith, and Iaria (2021) show that consideration set heterogeneity can be characterized as an individual-specific fixed effect in panel data when preferences are logit. Assuming that choice sets are either stable over time (with panel data) or across individuals (with cross-sectional data), preferences can be recovered from choice probabilities. In a setting where individuals only have capacity to consider a certain number of alternatives, Dardanoni et al. (2020) show that consideration probabilities can be

6. Davis and Schiraldi (2014) provide generalizations of multinomial logit models that permit asymmetries, but they explicitly note that these models cannot be rationalized by an underlying random utility interpretation and do not attempt to use these asymmetries to identify inattention.

identified in a setting with homogeneous preferences from a single cross-section of aggregate choice shares.

In this article, we do not impose a particular functional form on the nature of preference heterogeneity in the population. Thus, our results do not rely on the logit functional form and encompass all of the standard functional-form assumptions made on preference heterogeneity. In contrast to Crawford, Griffith, and Iaria (2021), our identification result relies on the insight that changes in product characteristics alter the probability that a consumer pays attention to a particular set of products and thus unobserved choice sets can vary over time and across markets, and our result does not require panel data.

III. Model and Identification

In this section, we show that assumptions commonly made in discrete-choice models are sufficient for identification in several standard models of imperfect consideration. Our central insight is that violations of Slutsky symmetry constructively identify the probability that consumers consider various subsets of products.

III.A. Basic Framework

We consider an individual i who makes a discrete choice among $J+1$ products, $\mathcal{J}=\{0,1,\ldots,J\}$, with $J \geqslant 1$. Each product j is associated with a price, p_j. The price vector $\mathbf{p}=[p_0,\ldots,p_J]$ is supported on $\mathbb{R}^{J+1}_{++}$. Our framework naturally incorporates additional characteristics ($\mathbf{x}_j$), consumer microdata ($\mathbf{z}_i$), and interactions between consumer and product characteristics. However, variation in these additional characteristics is not required for our identification result, and thus we suppress the dependence of choice on $\mathbf{x}_j$ and $\mathbf{z}_i$ in what follows. Our identification argument focuses on price variation, although it extends to variation in any attribute satisfying the assumptions we state below.

The (unobserved) set of goods that a consumer considers is called the consideration set. We first present a general consideration set model to describe the relationship between asymmetries and imperfect consideration before presenting the DSC, ASC, and hybrid models. Let $\mathcal{P}(\mathcal{J})$ represent the power set of goods, with any given element of $\mathcal{P}(\mathcal{J})$ indexed by C. The set of consideration sets containing good j is then given as:

$$\mathbb{P}(j)=\{C:\{0,j\}\subseteq C\in\mathcal{P}(\mathcal{J})\}. \tag{1}$$

In all of the models we investigate, observed choice probabilities take the following form:

$$s_j(\mathbf{p}) = \sum_{C \in \mathbb{P}(j)} \pi_C(\mathbf{p}) s_j^\star(\mathbf{p}|C), \tag{2}$$

where $s_j \equiv s_j(\mathbf{p})$ is the observed probability of good j being bought given market prices $\mathbf{p}$, $\pi_C(\mathbf{p})$ gives the probability that the set of goods C is considered given observable characteristics, and $s_j^\star(\mathbf{p}|C)$ gives the probability that good j is chosen from the consideration set C. As $\pi_C(\mathbf{p})$ and $s_j^\star(\mathbf{p}|C)$ represent proper probabilities, we have:

$$\sum_{C \in \mathcal{P}(\mathcal{J})} \pi_C(\mathbf{p}) = 1, \quad \sum_{j \in C} s_j^\star(\mathbf{p}|C) = 1. \tag{3}$$

In this article, the structural objects of interest are the consideration set probabilities, $\pi_C(\mathbf{p})$, and the unobserved latent choice probabilities, $s_j^\star(\mathbf{p}|C)$. We do not directly address the identification of preference parameters given knowledge of $s_j^\star(\mathbf{p}|C)$ nor the identification of, for example, search costs given consideration probabilities in any generality. The parameters of any utility model that are identified from choice behavior with full consideration, and the parameters of models that provide microfoundations for consideration sets given consideration probabilities will follow from our identification results. Our aim is to provide general identification results that can be tailored by applied researchers to special cases of the framework considered here.

1. Baseline Theory Assumptions. We assume that choice probabilities satisfy the standard Daly-Zachary conditions within a consideration set (Daly and Zachary 1978), notably cross-derivative symmetry and an absence of nominal illusion.[7]

ASSUMPTION 1. *Daly-Zachary Conditions:* unobserved latent choice probabilities, $s_j^\star(\mathbf{p}|C)$, satisfy the following conditions everywhere on $\mathbb{R}_{++}^{J+1}$:

7. See Anderson, De Palma, and Thisse (1992) and Koning and Ridder (2003) for further discussion of these conditions.

i. PROPERTIES: $s^\star_j(\mathbf{p}|C) \geqslant 0$, $\sum_{j\in C} s^\star_j(\mathbf{p}|C) = 1$, and

$$\frac{\partial^J s^\star_j(\mathbf{p}|C)}{\partial p_0 \ldots \partial p_{j-1} \partial p_{j+1} \ldots \partial p_J}$$

exists finitely, is $\geqslant 0$, and is continuous.[8]

ii. SYMMETRY: cross-price derivatives are symmetric:

$$\frac{\partial s^\star_j(\mathbf{p}|C)}{\partial p_{j'}} = \frac{\partial s^\star_{j'}(\mathbf{p}|C)}{\partial p_j}.$$

iii. ABSENCE OF NOMINAL ILLUSION:

$$s^\star_j(\mathbf{p} + \delta|C) = s^\star_j(\mathbf{p}|C).$$

Standard additive random utility models (ARUMs) imply choice probabilities that satisfy these conditions, including the nested and mixed logit models. For example, the following indirect utility function is consistent with Assumption 1:

$$u_j = v(p_j) + \epsilon_j \tag{4}$$

$$= \alpha_j - \beta p_j + \epsilon_j, \tag{5}$$

where $\epsilon_j \perp p_{j'}$ for all $j, j' \in \mathcal{J}$ and the joint distribution of ϵ is absolutely continuous and nondefective.[9]

2. Baseline Data Assumptions. In discussing identification, we treat the probability of selecting good j conditional on observables $\mathbf{p}$, $s_j(\mathbf{p})$, and all cross-derivatives as known for all possible prices $\mathbf{p} \in \mathbb{R}^{J+1}_{++}$. This is standard when addressing nonparametric identification of structural functions (Berry and Haile 2016).

ASSUMPTION 2. *Population Market Shares, Own- and Cross-Price Derivatives Observed at* $\mathbf{p} \in \mathbb{R}^{J+1}_{++}$*:* the observables consist of

8. Intuitively, the derivative condition states that if all goods except j get more expensive, then the probability of choosing j should not decrease.

9. The key restriction that consumers value price equally across choices is substantive, although it is often theoretically well motivated.

the variables:

$$\left\{ s_j(\mathbf{p}),\ \frac{\partial s_j(\mathbf{p})}{\partial p_{j'}} \right\}_{j,j' \in \mathcal{J}}. \tag{6}$$

Loosely, we are considering a scenario in which we have enough markets (across which prices vary) and individuals in these markets such that choice probabilities and their derivatives conditional on observables can be nonparametrically estimated. However, we do not know the extent to which observed choice probabilities reflect consideration versus preferences. In practice, one rarely has enough data to nonparametrically estimate $s_j(\mathbf{p})$; the purpose of our identification proof is to show that practically necessary functional-form restrictions are not required for identification (following Berry and Haile 2014). Prior work on semiparametric identification of multinomial choice models without consideration sets has assumed large price support (Lewbel 2000; Matzkin 2007). This assumption therefore provides a natural benchmark for exploring identification under ideal conditions. We discuss in the text and Online Appendix A where limited price variation will suffice.

3. Consideration Set Framework. Given Assumption 1, in our baseline model only one mechanism is available to generate cross-derivative asymmetries: imperfect consideration.[10] Theorem 1 makes this point formally.

THEOREM 1. ASYMMETRIES AND NOMINAL ILLUSION IMPLY IMPERFECT CONSIDERATION. Let observed choice probabilities have the following structure:

$$s_j(\mathbf{p}) = \sum_{C \in \mathbb{P}(j)} \pi_C(\mathbf{p}) s_j^{\star}(\mathbf{p}|C). \tag{7}$$

10. Not all models of inattention generate cross-price asymmetries. For example, Matejka and McKay (2014) show that when actions are homogeneous a priori and exchangable in the decision maker's prior, and the information strategy is time invariant, a rational inattention model provides a foundation for the multinomial logit (which yields symmetric cross-derivatives). Models in which consideration is independent of product characteristics would also have observed choice probabilities that satisfy the Daly-Zachary conditions, provided that the latent choice probabilities satisfy those conditions.

Given Assumption 1 and assuming that $\pi_C(\mathbf{p})$ is differentiable with respect to $\mathbf{p}$, if

$$\frac{\partial s_j(\mathbf{p})}{\partial p_{j'}} \neq \frac{\partial s_{j'}(\mathbf{p})}{\partial p_j} \tag{8}$$

or

$$s_j(\mathbf{p}) \neq s_j(\mathbf{p}+\delta) \tag{9}$$

for $\delta \neq 0$, then $\pi_{\mathcal{J}}(\mathbf{p}) < 1$, where $\pi_{\mathcal{J}}(\mathbf{p})$ is the probability that an individual considers all goods $\mathcal{J} = \{0, \ldots, J\}$. The proof is in Online Appendix A.

To make progress toward point identification of the structural functions of interest, we must place some additional restrictions on consideration set probabilities. If $\pi_C(\mathbf{p})$ are allowed to vary arbitrarily, then point identification of the underlying structural functions is not possible without additional information on what consumers considered (Manzini and Mariotti 2014).[11]

We derive identification results for the two most common consideration set models found in the applied literature, the DSC model and the ASC model, and a hybrid model that combines these two approaches.[12] The DSC model assumes the existence of an inside default good and allows the probability of considering all alternative options to vary only as a function of the characteristics of that default (Moshkin and Shachar 2002; Ho, Hogan, and Scott Morton 2017; Hortaçsu, Madanizadeh, and Puller 2017;

11. Several papers in the literature produce partial identification results in more general cases, such as Masatlioglu, Nakajima, and Ozbay (2012), Cattaneo et al. (2020), and Barseghyan et al. (2021).

12. Compared with some fully microfounded models of inattention, the DSC and ASC models can permit more general patterns of behavior. For example, a rational inattention model imposes that product attributes should affect attention in proportion to their value, but this need not be the case. However, our agnosticism means that we cannot identify how out-of-sample variation will affect attention, and our approach alone does not facilitate the identification of search costs without further structure. One could use the search probabilities recovered by our model to help identify a structural model that would identify these costs. In Section V, we highlight these points and evaluate the robustness of our normative evaluation to alternative assumptions about these unknown values.

Heiss et al. 2016).[13] The ASC model permits each good to have an independent probability of being considered that depends on the characteristics of that good (Goeree 2008; Manzini and Mariotti 2014; Kawaguchi, Uetake, and Watanabe forthcoming) and has been a popular model in marketing for many years (Swait and Ben-Akiva 1987; Ben-Akiva and Boccara 1995; Van Nierop et al. 2010).[14]

The ASC and DSC models impose substantive restrictions on the data (even when combined into a hybrid model). First, both models impose that the unobservable determinants of attention and utility are uncorrelated. This restriction can be relaxed but not without additional instruments (see Online Appendix A.7). Second, neither model allows for the possibility of correlated unobservable shocks to attention probabilities. Third, the models require at least one restriction on how attributes of goods are allowed to perturb attention probabilities for rival goods. These restrictions are not without loss, and their plausibility must be assessed in a context-specific way. If interest lies in scenarios that cannot be nested in this hybrid framework, in Online Appendix A we show in a more general environment that features of consideration probabilities are identified using cross-derivative asymmetries.

III.B. The Default Specific Model

Under the DSC model, the market shares of the default (good 0) and nondefault goods take the form:

$$\begin{aligned} s_0(\mathbf{p}) &= (1-\mu(p_0)) + \mu(p_0)s_0^\star(\mathbf{p}|\mathcal{J}) \\ s_j(\mathbf{p}) &= \mu(p_0)s_j^\star(\mathbf{p}|\mathcal{J}) \quad \text{for } j>0, \end{aligned} \tag{10}$$

where the differentiable function $\mu(p_0)$ gives the probability of considering all available products, while $s_j^\star(\mathbf{p}|\mathcal{J})$ gives choice probability for good $j \in \mathcal{J}$ conditional on considering all products.

Note that for simplicity, in the main text we assume:

13. This model can be straightforwardly microfounded in a rational inattention framework: only if the characteristics of the default get sufficiently bad do consumers pay a cost to search among all available products.

14. The ASC model is also supported by direct empirical evidence: Aguiar et al. (2018) conduct a lab experiment in which they observe choices from every possible subset of products and can thus recover flexible consideration set probabilities, finding that consideration patterns in the data can be rationalized by the ASC model with independent choice probabilities.

i. A homogeneous default good; this is to avoid introducing an i subscript. In Online Appendix A, we show that our results extend without complication to the case with heterogeneous defaults across consumers.

ii. That $\mu(\cdot)$ is a function of the characteristics of the default good only; this is to follow the existing literature. We show in Online Appendix A that our results extend without complication to the case where $\mu(\cdot)$ is also a function of the characteristics of any strict subset of nondefault goods.[15]

1. Identifying Changes in Consideration Probabilities. The key to our identification argument is that full consideration implies symmetric cross-price derivatives in standard discrete-choice models (Slutsky symmetry). However, with imperfect consideration, cross-derivative asymmetries arise. Differentiating equation (10) and using the fact that the market shares conditional on full consideration satisfy symmetry, we obtain:

$$\frac{\partial s_j(\mathbf{p})}{\partial p_0} - \frac{\partial s_0(\mathbf{p})}{\partial p_j} = \mu_0 \frac{\partial s_j^\star(\mathbf{p}|\mathcal{J})}{\partial p_0} + \frac{\partial \mu_0}{\partial p_0} s_j^\star(\mathbf{p}|\mathcal{J}) - \mu_0 \frac{\partial s_0^\star(\mathbf{p}|\mathcal{J})}{\partial p_j} \tag{11}$$

$$= \frac{\partial \mu_0}{\partial p_0} s_j^\star(\mathbf{p}|\mathcal{J}) \tag{12}$$

$$= \frac{\partial \log(\mu_0)}{\partial p_0} s_j(\mathbf{p}), \tag{13}$$

where $\mu_0 \equiv \mu(p_0)$ and the last line uses the fact that $s_j^\star(\mathbf{p}|\mathcal{J}) = \frac{s_j(\mathbf{p})}{\mu_0}$. Thus, changes in the probability of full consideration are directly identified from data on choice probabilities:

$$\frac{\partial \log(\mu_0)}{\partial p_0} = \frac{1}{s_j(\mathbf{p})} \left[\frac{\partial s_j(\mathbf{p})}{\partial p_0} - \frac{\partial s_0(\mathbf{p})}{\partial p_j} \right]. \tag{14}$$

Intuitively, if the price of the default plan perturbs consideration by causing consumers to "wake up" (the left-hand side), then the nondefault plan will be more sensitive to the price of the

15. See Online Appendix A.6.

default plan than is the default plan to the price of the nondefault plan. This is a behavioral pattern noted in the marketing and health insurance literature by Ho, Hogan, and Scott Morton (2017) and Moshkin and Shachar (2002).

THEOREM 2. IDENTIFICATION OF CHANGES IN DSC CONSIDERATION PROBABILITIES. Given Assumptions 1 and 2, then $\frac{\partial \log(\mu_0)}{\partial p_0}$ is constructively identified.

2. Identifying the Level of Consideration. In recovering the derivative of the log consideration probability, μ_0 is identified up to a constant by integrating over the support of p_0:

$$\log(\mu(\infty)) - \log(\mu(\tilde{p}_0)) = \int_{\tilde{p}_0}^{\infty} \frac{1}{s_j(\mathbf{p})}\left[\frac{\partial s_j(\mathbf{p})}{\partial p_0} - \frac{\partial s_0(\mathbf{p})}{\partial p_j}\right] \mathrm{d}p_0. \tag{15}$$

Identifying the level of consideration (and thus latent market shares) requires an additional assumption to pin down the constant of integration. Assuming that consumers are prompted to consider good j when p_0 reaches an extreme value enables the level of consideration to be identified (i.e., $\log(\mu(\infty)) = 0$ in equation (15)). This assumption is analogous to those made in the literature on nonparametric identification of multinomial discrete choice models (Lewbel 2000; Berry and Haile 2009), treatment effects (Heckman and Vytlacil 2005; Lewbel 2007; Magnac and Maurin 2007), the identification of binary games and entry models (Tamer 2003; Fox, Hsu, and Yang 2012; Lewbel and Tang 2015), and the use of special regressors more generally. This assumption is testable in our setting by checking that cross-derivative differences are symmetric at high default prices.

ASSUMPTION DSC. As $p_0 \to \infty$, $\mu(p_0) \to 1$.

THEOREM 3. IDENTIFICATION OF $\mu(p_0)$ IN THE DSC MODEL. Given Assumptions 1, 2, and DSC, then consideration probabilities are constructively identified as:

$$\mu(\tilde{p}_0) = \exp\left(-\int_{\tilde{p}_0}^{\infty} \frac{1}{s_j(\mathbf{p})}\left[\frac{\partial s_j(\mathbf{p})}{\partial p_0} - \frac{\partial s_0(\mathbf{p})}{\partial p_j}\right] dp_0\right). \tag{16}$$

Note that nonparametric identification of $\mu(p_0)$ requires substantially less observed price variation than that implied by

Assumption 2.[16] Furthermore, commonly employed functional-form assumptions on consideration probabilities substantially reduce the amount of price variation required to identify the level of consideration,[17] even when no further parametric assumptions are placed on preferences. For example, let consideration take the following form:[18]

$$\mu(p_0) = \frac{\exp(\gamma_0 + \gamma_p p_0)}{1 + \exp(\gamma_0 + \gamma_p p_0)}. \tag{17}$$

In this scenario, all that is required is for there to exist at least two levels of the default price at which market shares and cross-derivatives are observed.

THEOREM 4. IDENTIFICATION OF $\mu(p_0)$ WITH LOGIT CONSIDERATION IN THE DSC MODEL. Given Assumption 1 and two strictly positive price vectors $\{\mathbf{p}^a, \mathbf{p}^b\}$ with $p_0^a \neq p_0^b$ at which market shares and cross-price derivatives are observed, then $[\gamma_0, \gamma_p]$ are identified where

$$\mu(p_0) = \frac{\exp(\gamma_0 + \gamma_p p_0)}{1 + \exp(\gamma_0 + \gamma_p p_0)}. \tag{18}$$

The proof is in Online Appendix A.

III.C. The Alternative Specific Model

Under the ASC approach, consideration set probabilities take the form:

$$\pi_C(\mathbf{p}) = \prod_{j \in C} \phi_j(p_j) \prod_{j' \notin C} (1 - \phi_{j'}(p_{j'})), \tag{19}$$

where the probability of good j being considered, $\phi_j \equiv \phi_j(p_j)$, is a differentiable function of own characteristics only and $\phi_j(p_0) = 1$

16. We require that there exists for each $p_0 \in \mathbb{R}_{++}$ a good j and price vector $\mathbf{p}_{\neg 0} = \{p_k\}_{k \in \mathcal{J}/0}$ such that $\left\{s_j(p_0, \mathbf{p}_{\neg 0}), \frac{\partial s_j(p_0, \mathbf{p}_{\neg 0})}{\partial p_0}, \frac{\partial s_0(p_0, \mathbf{p}_{\neg 0})}{\partial p_j}\right\}$ exist and are observed.

17. For examples of papers using this functional-form assumption seee Heiss et al. (2016) for the DSC model and Goeree (2008) for the ASC model.

18. This is the assumption we will be making in our empirical applications.

for all p_0. Observed market shares then take the form:

$$s_j(\mathbf{p}) = \sum_{C\in\mathbb{P}(j)} \prod_{l\in C} \phi_l(p_l) \prod_{l'\notin C} \big(1 - \phi_{l'}(p_{l'})\big)\, s_j^{\star}(\mathbf{p}|C). \tag{20}$$

Even in this more complicated model, changes in consideration probabilities can be expressed as a function of observable differences in cross-derivatives and market shares. This is despite the fact that the probability of considering a particular set of goods is a function of the characteristics of all products in the market, although in the manner constrained by the theoretical framework.

1. Identifying Consideration Probabilities. Even in this richer setting, changes in consideration probabilities can be expressed as a linear function of observables. Let $\bar{\mathbf{p}}_j$ give the price vector $\mathbf{p}$ under which p_j is approaching ∞ with $s_j(\bar{\mathbf{p}}_j) = 0$.[19] As shown fully in Online Appendix A, one can express cross-derivative differences between default and nondefault products as:

$$\frac{\partial s_0(\mathbf{p})}{\partial p_j} - \frac{\partial s_j(\mathbf{p})}{\partial p_0} = \frac{\partial \log(\phi_j)}{\partial p_j}(s_0(\mathbf{p}) - s_0(\bar{\mathbf{p}}_j)), \tag{21}$$

while cross-derivative differences for $j, j' \neq 0$ are:

$$\begin{aligned}\frac{\partial s_j(\mathbf{p})}{\partial p_{j'}} - \frac{\partial s_{j'}(\mathbf{p})}{\partial p_j} &= \frac{\partial \log(\phi_{j'})}{\partial p_{j'}}(s_j(\mathbf{p}) - s_j(\bar{\mathbf{p}}_{j'})) \\ &\quad - \frac{\partial \log(\phi_j)}{\partial p_j}(s_{j'}(\mathbf{p}) - s_{j'}(\bar{\mathbf{p}}_j)).\end{aligned} \tag{22}$$

Equations (21) and (22) relate unobservable changes in consideration probabilities to observed cross-derivative differences and market shares.

For simplicity, in the main text we provide the just-identified conditions for an inside default. In Online Appendix A, we give identification results based on the full system of cross-derivative differences defined by equations (21) and (22), which are also suitable for scenarios where the default is the outside good. Analogous arguments to those made with respect to the DSC model can be used to prove identification of the level of $\phi_j(\cdot)$. Without

19. Please see Online Appendix A for a formal discussion of $\bar{\mathbf{p}}_j$.

assuming a particular functional-form assumption for $\phi_j(\cdot)$, one must continue to rely on large-support assumptions for nonparametric identification, namely, that consumers are prompted to consider a product with probability one at extreme values of the covariates (Assumption ASC.i).[20] These data requirements are, however, reduced when one assumes a parametric form for consideration probabilities. Assumption ASC.ii imposes that there is some substitution to good-0 when the price of nondefault goods is high; a weak assumption that is easily tested.

ASSUMPTION ASC.I: As $p_j \to \infty$, $\phi_j(p_j) \to 1$.

ASSUMPTION ASC.II: $s_0(\mathbf{p}) - s_0(\bar{\mathbf{p}}_j) \neq 0$ at all $\mathbf{p} \in \mathbb{R}^{J+1}_{++}$.

THEOREM 5. IDENTIFICATION OF $\phi_j(p_j)$ IN THE ASC MODEL. Given Assumptions 1, 2, ASC.i, and ASC.ii, then $\phi_j(p_j)$ for $j = 1, \ldots, J$ are identified at all $\mathbf{p} \in \mathbb{R}^{J+1}_{++}$:

$$\tag{23} \begin{gathered} \frac{\partial \log(\phi_j)}{\partial p_j} = \frac{\frac{\partial s_0(\mathbf{p})}{\partial p_j} - \frac{\partial s_j(\mathbf{p})}{\partial p_0}}{s_0(\mathbf{p}) - s_0(\bar{\mathbf{p}})} \\ \phi_j(\tilde{p}_j) = \exp\left(-\int_{\tilde{p}_j}^{\infty} \frac{\frac{\partial s_0(\mathbf{p})}{\partial p_j} - \frac{\partial s_j(\mathbf{p})}{\partial p_0}}{s_0(\mathbf{p}) - s_0(\bar{\mathbf{p}}_j)} \, \mathrm{d}p_j \right). \end{gathered}$$

III.D. The Hybrid Consideration Set Model

The assumptions made to identify consideration probabilities in the ASC model are also sufficient to identify a hybrid model that combines the ASC and DSC models. Combining the models seems natural in many applied settings. For example, with many options, it seems implausible that consumers would consider all goods in the feasible choice set on waking up. However, conditions for the identification of this combined hybrid framework have not been previously investigated nor has the framework been harnessed for applied research until this article.

20. Note that this is to pin down the constant of integration. An alternative assumption that locates the level of consideration might be more natural in some applications (e.g., consider a good if the level of an attribute falls to a particularly low level).

Under the hybrid consideration set model, the market shares of the default and nondefault goods take the form:

$$s_0(\mathbf{p}) = (1-\mu(p_0)) + \mu(p_0)\sum_{C\in\mathbb{P}(0)}\prod_{l\in C}\phi_l(p_l)\prod_{l'\notin C}(1-\phi_{l'}(p_{l'}))s_0^{\star}(\mathbf{p}|C)$$

$$s_j(\mathbf{p}) = \mu(p_0)\sum_{C\in\mathbb{P}(j)}\prod_{l\in C}\phi_l(p_l)\prod_{l'\notin C}(1-\phi_{l'}(p_{l'}))s_j^{\star}(\mathbf{p}|C) \quad \text{for } j>0,$$

(24)

where $\phi_0(p_0) = 1$ for all $p_0 \in \mathbb{R}_{++}$. Restricting $\phi_j(p_j) = 1$ for all $j > 0$ gives the DSC model. Restricting $\mu(p_0) = 1$ gives the ASC model.

In this model, cross-derivative differences between nondefault goods have the same structure as in the ASC model (equation (22)). However, cross-derivatives involving the default good take a slightly different form because of the effect of the default good on the probability of waking up:

$$\text{(25)} \quad \frac{\partial s_j(\mathbf{p})}{\partial p_0} - \frac{\partial s_0(\mathbf{p})}{\partial p_j} = \frac{\partial \log(\mu_0)}{\partial p_0}s_j(\mathbf{p}) - \frac{\partial \log(\phi_j)}{\partial p_j}(s_0(\mathbf{p}) - s_0(\bar{\mathbf{p}}_j)).$$

Let the system of equations defined by equations (22) and (25) be expressed as:

$$\text{(26)} \quad c(\mathbf{p}) = D(\mathbf{p})A(\mathbf{p}),$$

where $c(\mathbf{p})$ is a $\frac{1}{2}J(J+1)$-vector of cross-derivative differences at prices $\mathbf{p}$, $D(\mathbf{p})$ is the coefficient matrix of market share differences, and $A(\mathbf{p}) = \left[\frac{\partial \log \mu_0}{\partial p_0}, \frac{\partial \log \phi_1}{\partial p_1}, \ldots, \frac{\partial \log \phi_J}{\partial p_J}\right]$ is the $J+1$-vector of log consideration probability derivatives.[21] Because there are typically more than $J+1$ cross-derivative differences, it is convenient to work with the system:[22]

$$\text{(27)} \quad D(\mathbf{p})'c(\mathbf{p}) = D(\mathbf{p})'D(\mathbf{p})A(\mathbf{p}).$$

If $D(\mathbf{p})'D(\mathbf{p})$ is full rank, there is a unique solution to this system and changes in consideration probabilities are uniquely identified from choice data. Online Appendix A discusses the restrictions on structural functions required for this rank condition

21. See Online Appendix A for illustrations of the structure of these matrices.
22. Alternative weighting matricies, W_m, can be used: $D_m' W_m D_m$.

to hold. Intuitively, goods being imperfect substitutes and being considered with strictly positive probability at all price vectors will suffice. A strength of our approach is that the rank condition is testable given market share data. When it fails, changes in consideration probabilities can be recovered at price vectors where the assumption holds.

ASSUMPTION HYBRID. *Rank Condition*: The matrix $D(\mathbf{p})'D(\mathbf{p})$ is full rank at all $\mathbf{p} \in \mathbb{R}_{++}^{J+1}$.

THEOREM 6. (IDENTIFICATION OF CONSIDERATION PROBABILITIES IN THE HYBRID CONSIDERATION SET MODEL) Given Assumptions 1, 2, DSC, ASC.i, ASC.ii, and Hybrid, then $\mu(p_0)$ and $\phi_j(p_j)$ are identified at all $\mathbf{p} \in \mathbb{R}_{++}^{J+1}$. The proof is in Online Appendix A.

III.E. Identifying Latent Market Shares

We have focused our attention so far on the identification of consideration probabilities. In the DSC model, identification of latent market shares is trivial once $\mu(p_0)$ is known as $s_j^\star(\mathbf{p}) = \frac{\mu(p_0)}{s_j(\mathbf{p})}$. In the ASC and Hybrid frameworks, however, matters are more complicated given that there exist 2^J independent latent choice probabilities for any given price vector.[23]

In Online Appendix A, we show how the restrictions deriving from "nominal illusion" facilitate the identification of the 2^J independent latent choice probabilities in the ASC and hybrid models, $s_j^\star(\mathbf{p}|C)$. To provide intuition for our result, imagine a rise in all prices by some amount δ such that relative prices remain unchanged. Given the Daly-Zachary conditions, this price shift can change consideration probabilities but does not alter latent choice probabilities conditional on consideration. Thus, observed market shares can vary even though nominal illusion would suggest invariance of choices to price changes that do not alter relative prices. Hastings and Shapiro (2013) establish this behavioral pattern for gasoline choice. In Online Appendix A, we show how this variation is sufficient to identify latent choice probabilities.

23. This identification problem is analogous to the problem of identifying the long regression. While the functions of interest are typically only partially identified without instruments (Henry, Kitamura, and Salanié 2014), we show that optimizing behavior here results in point identification of the objects of interest.

III.F. Overidentification

Given our assumptions, imperfect consideration is the only mechanism giving rise to asymmetric cross-derivatives. Relaxing our background assumptions might, however, give rise to alternative sources of asymmetry that our framework could incorrectly attribute to inattention. We note that our model is overidentified, providing the potential to test the validity of the consideration set model outlined in this article, and that the asymmetries predicted by our framework have a particular structure.[24] With $J > 2$, the derivative of the log of consideration probabilities are overidentified; intuitively there are more cross-price derivative differences than consideration probabilities. In the hybrid model, for example, there are

$$\underbrace{\frac{1}{2}J(J+1)}_{\text{\# Independent Cross-Deriv. Diffs}} \qquad - \quad \overbrace{(J+1)}^{\#\phi_j \text{ Derivatives}} \tag{28}$$

overidentifying restrictions for changes in consideration probabilities.[25]

IV. Validation and Estimation

In this section, we discuss estimation of our model and validate our approach. To do so, we must observe "true" consideration probabilities; something that is very rare in observational data. We thus validate the practical relevance of our identification result in the lab. We ask consumers to make choices from known subsets of 10 goods that are generated according to the ASC model. We ask whether we can recover the (known) consideration probabilities as well as preferences conditional on consideration using information only on observed choices. This test goes beyond a simulation exercise by showing that we can use our model to recover consideration probabilities in a setting where our functional-form assumptions on latent choice probabilities (i.e., preferences) are not guaranteed to hold.

24. This also facilitates testing against spurious asymmetries that might arise due to data issues, such as measurement error.

25. Similar reasoning shows that latent market shares are overidentified when there are more shifts in prices than latent market shares.

IV.A. Set Up

To experimentally validate our approach, we conducted a discrete-choice consumption experiment with 149 Yale students.[26] We selected 10 goods sold at the Yale bookstore with list prices ranging from $19.98 to $24.98. These goods and their list prices are shown in Table 7 in Online Appendix B.5. Each participant was endowed with $25 and made 50 choices from randomly chosen subsets of the 10 goods with randomized prices (one-third of the list price plus a uniformly distributed amount between $0 and $16). A sample product selection screen is shown in Figure I. Consumers were shown images of all the products in the displayed subset along with the (randomly chosen) prices and asked to select their preferred option. After making all 50 choices, one of these choices was randomly selected and they were given that item as well as $25 minus the price of the item in cash.[27] In total, we ran the experiment with 149 participants, resulting in 7,450 choices.[28]

We treat the subset of products that appear on a respondent's screen as the consideration set.[29] The probability that each good appeared on the screen was fixed by us in advance—this probability varied across goods and with prices such that goods were more likely to be considered (i.e., appear on a respondent's screen) if they had a higher price (perhaps mimicking the behavior of a retailer who places their highest-margin products where they are

26. There were 150 participants in total, but one participant's data was not recorded properly because they refreshed the browser several times during the experiment, which interfered with data capture; this participant is dropped from the final analysis.

27. Participants had to chose one item and thus could not simply take the $25 payment. Although this does not introduce a bias to our results because participants' ranking over the remaining options are unaffected, we do not think that many of our participants would have simply taken the cash if given the chance. First, no participants chose the lowest priced item in every round. Second, prices in the experiment were typically lower than the list price, creating arbitrage opportunities.

28. Prior to the experiment, participants were given several examples to illustrate the incentive scheme and were quizzed on their understanding. Seventy percent correctly answered our test of understanding (and all participants were told why their answer was correct or incorrect). Online Appendix Table 8 reports results using only this subset of users who passed this test and shows that results are qualitatively unchanged.

29. To increase the likelihood that participants considered all of the products they were presented with, we required them to spend at least 10 seconds looking at the screen before finalizing their choices.

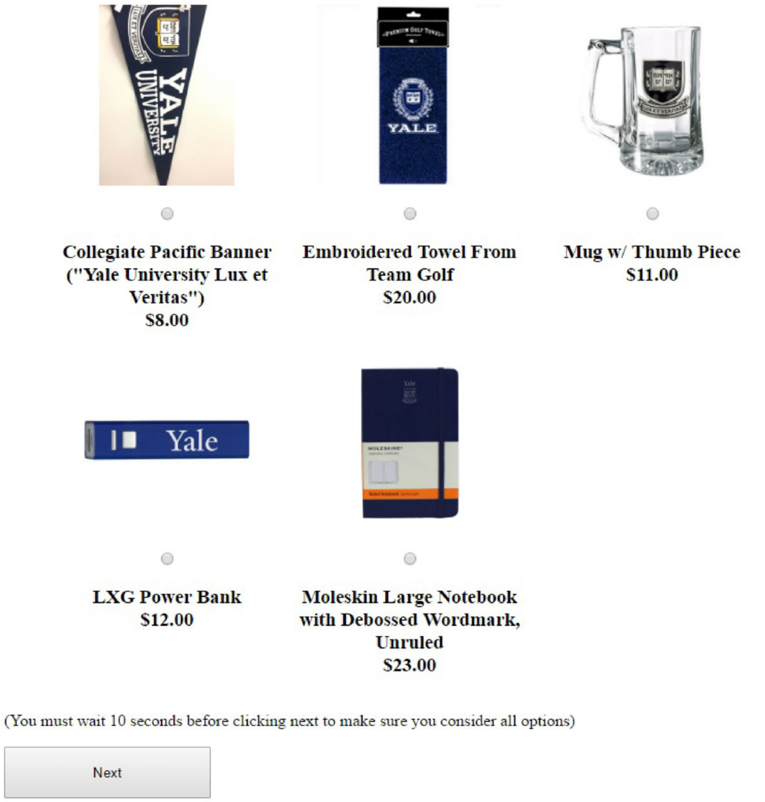

FIGURE I

Lab Experiment: Sample Product Selection Screen

most likely to be noticed).[30] We specified the probability that good j was in a participant i's round r consideration set as:[31]

$$\phi_{ijr} = Pr(\gamma_j + p_{irj}\gamma_p - \eta_{irj} > 0) \tag{29}$$

$$= \frac{\exp(\gamma_j + p_{irj}\gamma_p)}{1 + \exp(\gamma_j + p_{irj}\gamma_p)}, \tag{30}$$

30. Closing the model requires us to specify a good that is chosen if the consideration set is empty. We specify this as good 10. At the estimated parameter values, an empty consideration set has a 0.2% chance of occurring, so the choice of default does not affect estimation.

31. This is a similar empirical specification to that applied in the ASC literature to date (Goeree 2008).

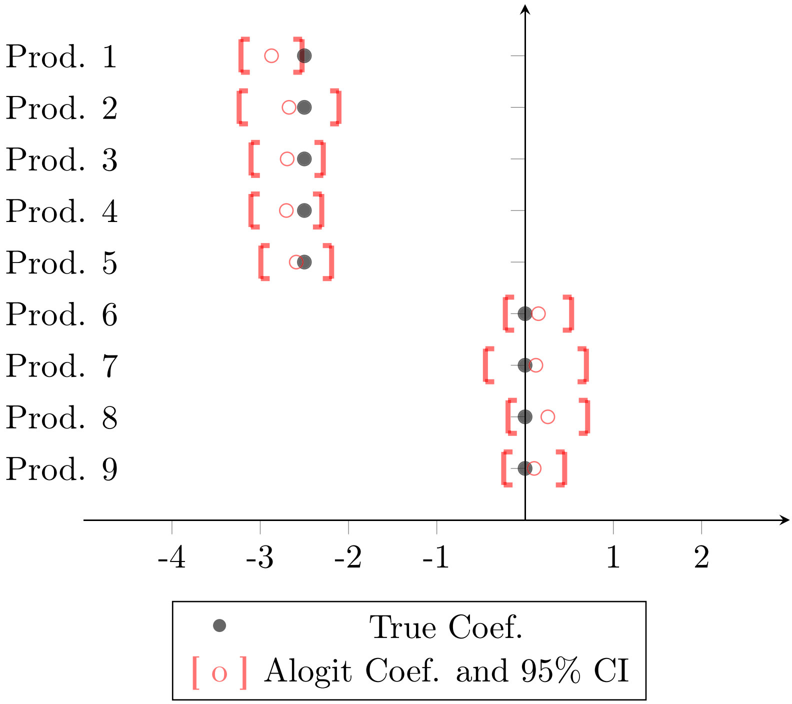

FIGURE II

Product Fixed Effects in Attention: Truth versus ASC Model

where η_{irj} is distributed logistic, p_{irj} gives the product's price, and γ_j is a product-specific fixed effect. The coefficients were chosen so that most choice sets would include two to seven products. See Table I for the precise coefficients. Our key question is whether we can accurately recover these parameters by harnessing asymmetries in cross-price responses.

IV.B. Estimation

Existing applications of consideration set models are typically estimated by maximum likelihood (Goeree 2008).[32] For validation purposes, the principal downside of this approach is the lack of transparency regarding what variation is driving our

32. Please see our STATA command alogit for estimation by maximum likelihood.

TABLE I
EXPERIMENTAL DATA ESTIMATION RESULTS

	Conditional logit (1)	ASC model: MLE (2)	ASC model: Indirect inf. (3)	Conditional on consideration (4)
Utility				
Price (dollars)	−0.054***	−0.196***	−0.157**	−0.173***
	(0.003)	(0.028)	(0.029)	(0.004)
Product 1	−1.411***	1.465***	0.613	0.368***
	(0.054)	(0.539)	(0.406)	(0.069)
Product 2	−1.955***	−0.065	0.133	−0.497***
	(0.069)	(0.478)	(0.573)	(0.080)
Product 3	−1.627***	0.625	0.344	0.093
	(0.059)	(0.476)	(0.403)	(0.073)
Product 4	−1.640***	0.629	0.443	0.088
	(0.060)	(0.466)	(0.386)	(0.073)
Product 5	−1.447***	0.707	0.934***	0.306***
	(0.056)	(0.478)	(0.354)	(0.070)
Product 6	−0.435***	−0.737***	−0.422	−0.581***
	(0.039)	(0.121)	(0.252)	(0.045)
Product 7	−0.855***	−1.280***	−0.474	−1.075***
	(0.045)	(0.141)	(0.234)	(0.051)
Product 8	−0.662***	−1.185***	−0.711***	−0.909***
	(0.041)	(0.137)	(0.271)	(0.048)
Product 9	−0.316***	−0.561***	−0.706***	−0.405***
	(0.038)	(0.118)	(0.184)	(0.044)
Attention				
Price (dollars)		0.137***	0.133***	0.15
		(0.017)	(0.027)	
Product 1		−2.872***	−2.636***	−2.5
		(0.177)	(0.359)	
Product 2		−2.674***	−2.914***	−2.5
		(0.288)	(0.544)	
Product 3		−2.695***	−2.599***	−2.5
		(0.209)	(0.360)	
Product 4		−2.704***	−2.703***	−2.5
		(0.205)	(0.389)	
Product 5		−2.592***	−3.042***	−2.5
		(0.204)	(0.389)	
Product 6		0.152	−0.464	0
		(0.192)	(0.367)	
Product 7		0.123	−0.925***	0
		(0.292)	(0.345)	
Product 8		0.258	0.476	0
		(0.230)	(0.425)	
Product 9		0.103	0.779	0
		(0.176)	(0.532)	

Notes. This table reports coefficient estimates from conditional logit and the ASC model. Estimates are the coefficients in the utility and attention equations (not marginal effects). The conditional logit coefficients are recovered from estimating a model assuming all 10 possible goods are considered. The "conditional on consideration" utility parameters are estimated using a conditional logit model that conditions on the actual choice set consumers faced. The true attention parameters are set by us in advance. The ASC model also includes a constant in consideration probabilities. *** Denotes significance at the 1% level, ** significance at the 5% level, and * significance at the 10% level.

results. Are the estimated consideration probabilities driven by the asymmetries in the choice probabilities or by parametric assumptions made in specifying the model? To address this question directly, we estimate the model by indirect inference in addition to the more conventional maximum likelihood approach.[33]

1. Indirect Inference. Indirect inference involves specifying a flexible auxiliary model, estimating that model on the observational data, and then choosing structural parameters so that simulated data from the underlying structural model leads to the same auxiliary model estimates (Smith 1993; Gourieroux, Monfort, and Renault 1993). Following Keane and Smith (2003), we define a flexible auxiliary model characterized by the parameter vector θ. Our identification proof points to the importance of specifying an auxiliary model that permits asymmetric cross-elasticities.[34] We specify a flexible logit model, where we begin with a conventional logit model with good-specific price coefficients and then add additional interaction terms between the prices of alternative goods to capture asymmetries flexibly. That is, we specify the reduced-form auxiliary model for choice probabilities, $\widetilde{s}_{irj}$, as follows:

$$\widetilde{s}_{irj} = \frac{\exp(\widetilde{u}_{irj})}{\sum_k \exp(\widetilde{u}_{irk})} \tag{31}$$

$$\widetilde{u}_{irj} = \theta_j + \theta_j^0 p_{irj} + \sum_{j'} \theta_{jj'} p_{irj} p_{irj'}. \tag{32}$$

33. Our result is constructive, so nonparametric estimation of choice probabilities is possible in principle; in practice, the curse of dimensionality renders this approach infeasible in most applied settings of interest. A further consideration for applied researchers is that of endogeneity. We do not consider this in detail in this article given our focus on the identification issues arising from imperfect consideration alone. Goeree (2008) considers estimation of the ASC model in the presence of price endogeneity. More generally, our results show that if instruments can be used to identify the structural derivatives of choice probabilities with respect to product attributes conditional on any unobserved correlate of product attributes, then one can identify consideration probabilities. Estimation of these structural derivatives is nontrivial, and we intend to address it in future work.

34. Although consistent estimation does not require the auxiliary model to provide a correct statistical description of the observed data, if it does, then indirect inference has the same asymptotic efficiency as maximum likelihood. As discussed in Bruins et al. (2018), the specification of the auxiliary model should balance statistical and computational efficiency; one should choose an auxiliary model that is flexible enough to give a good description of the data, while also being relatively quick to estimate.

In our experimental setting with 10 goods, this specification gives rise to an auxiliary model with 119 parameters.

In our auxiliary model, the set of $\theta_{jj'}$ parameters capture asymmetric cross-derivatives. This specification has several desirable properties. We show in Online Appendix B that consideration set models with logit utility can be rewritten as a full-consideration model where the utility of each alternative j depends directly on the attributes of rival goods.[35] In the ASC case, we can derive our auxiliary model as a second-order Taylor expansion with respect to attributes of rival goods around the point where this dependence is zero (which yields the logit choice probabilities).[36] In addition, this specification nests the conventional logit model, which yields a symmetric substitution matrix, as a special case.

The representation result in Online Appendix B that motivates this auxiliary specification has a few other notable implications. The fact that consideration set models can be rewritten as random utility models where attributes of rival goods directly enter utility suggests a shortcoming of so-called BLP instruments, where the exclusion of rival characteristics from the utility of each good is relied on for identification (Berry, Levinsohn, and Pakes 1995). In addition, this representation shows that including fixed effects in a conventional model is not sufficient for consistent estimation given consideration sets.[37]

We estimate the auxiliary model using our experimental data to obtain parameter estimates $\hat{\theta}$:

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(y; p, \theta), \tag{33}$$

35. This is a similar insight to that pursued in Crawford, Griffith, and Iaria (2021).

36. Note that this provides one reason to prefer this auxiliary model to a flexible linear model. The flexible linear model is a Taylor expansion around a constant, whereas this model is a Taylor expansion around the logit choice probabilities.

37. In the related literature on choice-based sampling, the econometrician sees only a subset of goods from which consumers choose. In these models, one can sometimes consistently estimate preferences by controlling for alternative-specific constants (Manski and Lerman 1977; Bierlaire, Bolduc, and McFadden 2008). In our framework, this approach does not work because such constants cannot capture the direct dependence of utility on (variable) attributes of rival goods.

where

$$\mathcal{L}(y; p, \theta) = \sum_i \sum_r \sum_j 1(y_{ir} = j) \log \widetilde{s}_{irj}(p_{ir}; \theta), \tag{34}$$

where $y_{ir} \in \mathcal{J}$ records which option consumer i bought in round r of the experiment and y gives this choice variable stacked across rounds and individuals, and p gives the vector of prices, p_{irj}, stacked across individuals, rounds, and options.

Given prices p and structural parameters ψ, we use our consideration set model to simulate M statistically independent simulated data sets, $\{\widetilde{y}^m(\psi)\}^{m=1,\dots,M}$, by redrawing the structural error terms from their parametric distributions. In our empirical applications, we assume that the additive random error term in preferences, ϵ_{ijr}, is distributed i.i.d. Type 1 extreme value. Although this is (intentionally) a restrictive model of preferences, we ask if we are able to recover the process generating consideration sets even when preferences are modeled in this simplistic way. The auxiliary model is then estimated on each of the M simulated data sets to obtain a set of estimated parameter vectors $\widetilde{\theta}^m(\psi)$. Formally, $\widetilde{\theta}^m(\psi)$ solves:

$$\widetilde{\theta}^m(\psi) = \arg\max_\theta \mathcal{L}(\widetilde{y}^m(\psi); p, \theta), \tag{35}$$

where $\mathcal{L}(\cdot)$ is defined as in equation (34).

Indirect inference generates an estimate $\widehat{\psi}$ of the structural parameters that minimizes the distance between the parameters of the auxiliary model estimated on the observed and simulated data. Loosely speaking, the approach harnesses the insight that if one has the right data-generating process, operations performed on observed and simulated data should give the same answer. Let $\widetilde{\theta}(\psi) = M^{-1} \sum \widetilde{\theta}^m(\psi)$. Formally, $\widehat{\psi}$ solves:

$$\widehat{\psi} = \arg\min_\psi (\hat{\theta} - \widetilde{\theta}(\psi))' W (\hat{\theta} - \widetilde{\theta}(\psi)), \tag{36}$$

where W is a positive definite weighting matrix. Note that the set of structural errors used to generate the simulated data sets are held fixed for different values of ψ. As the sample size grows large, $\hat{\theta}$ and $\widetilde{\theta}(\psi)$ both converge to the same pseudo true value, θ_0, underlying the consistency of the approach (Gourieroux, Monfort, and Renault 1993).

IV.C. Results

Table I shows the results of our validation exercise. In column (4), we give the "true" consideration and preference coefficients. The true consideration coefficients are known with certainty given that they were specified by us (equation (30)). In the case of preference parameters, we take conditional logit parameters estimated by maximum likelihood using information on the actual choice sets that consumers faced as the relevant comparator, that is, we use all information on what respondents saw.[38]

Table I, columns (1)–(3) report parameter estimates that only use information about the product consumers actually chose and not information about the specific subset of 10 goods they could choose from in each instance. In Online Appendix B.5, we analyze the ability of the models to capture price elasticities. First, consider the maximum likelihood preference parameters shown in the top panel of Table I. The conditional logit model assuming a considered choice set of all 10 goods gives a price effect of −0.05, less than a third of the true value. This is because the conditional logit model wrongly infers from the fact that high-priced products are more likely to be considered (and thus chosen) that consumers do not really dislike high prices. Furthermore, the conditional logit fixed effects are systematically biased because they conflate attention and utility; products that are rarely in the considered choice set are assumed to be low utility.

In contrast, our consideration set models are able to accurately recover the process generating consideration and imply much more elastic price responses. We give results estimated by maximum likelihood and our indirect inference strategy. Both sets of results imply confidence intervals on the consideration price effect that include the true value. The consideration set models also recover preference fixed effects consistent with those estimated using all information on what products were considered. The intervals are relatively wide, but that is a feature, not a bug relative to the conditional logit model: the consideration set model correctly recognizes that rare products are rare and that only limited information is available about how much consumers value them.

38. We take a simple conditional-on-consideration conditional logit specification as the benchmark throughout; this is a natural generalization of the specification estimated in column (1), which uses no information on what items were shown to respondents. Online Appendix B.2 shows that this model fits the choice data well within consideration sets and serves as a good benchmark for comparison.

The consideration set model confidence intervals on the less rare products (products 6–9 in Table I) are reasonably precise.

Using our indirect inference strategy, we cannot reject the null hypothesis that the ASC model can explain the patterns reflected in the auxiliary model ($p = .4182$).[39] The restriction that consideration is independent of price is also rejected at all conventional confidence levels as is a full-consideration model with good-specific price coefficients (which thus allows for asymmetric cross-derivatives, although not in the manner predicted by limited consideration models). See Online Appendix B.5 for more details. These results provide strong evidence that our approach can discriminate between consideration and preferences using real choice data.

V. Limited Consideration and Smart Defaults in Medicare Part D

Disentangling the degree to which choice behavior reflects limited consideration versus preferences is important for market and policy design, especially in the case of insurance. A large literature finds that consumers choosing insurance plans fail to minimize costs and make systematic errors (Abaluck and Gruber 2011, 2016; Heiss et al. 2013; Bhargava, Loewenstein, and Sydnor 2015). Consumers are also highly inertial (Handel 2013); in Medicare Part D, where elderly consumers choose prescription drug insurance, over 90% of returning consumers choose the same plan as the previous year.

In response to such behavior, (Handel and Kolstad 2015) propose a "smart default" policy, under which consumers would be automatically assigned to lower-cost plans but given the option of switching back. Whether such a policy will make consumers better off depends on whether the associated cost savings outweigh the "acclimation costs" of learning to navigate a new plan.[40]

39. Online Appendix B gives the formal details of the goodness-of-fit test. In summary, the minimized value of the objective function is distributed chi-squared with degrees of freedom equal to the difference in the number of parameters in the auxiliary versus structural model. In Online Appendix B.5, we also report that an alternative full-consideration model with good-specific price parameters (which can generate cross-derivative asymmetries, although ones which have a different structure to a limited consideration model) is rejected at all conventional significance levels ($p = .0000$; $\chi^2 = 2{,}279$).

40. In Online Appendix C, we consider two alternative drivers of inertia. The first is spurious state dependence, wherein inertia arises because chosen plans are

Acclimation costs include costs such as scheduling deliveries for mail-order drugs or learning which of several chemically equivalent drugs are covered by any given plan. A smart default policy has not yet been implemented. In this article, we harness our framework to estimate limited consideration in Medicare Part D plan choice, allowing us to predict how consumers would respond to the policy and to normatively evaluate the results.[41]

1. Contribution of Our Approach. We model limited consideration using the hybrid version of our model. This allows consumers to exhibit two types of inattention: they can be asleep, in which case they remain enrolled in the plan they chose last year (the DSC model), or they can be awake and actively choose but attend to only a subset of options (the ASC model). This has important consequences for the evaluation of a smart default policy. In our model, full consideration implies that smart defaults will have no effect on behavior and acclimation costs must be huge to rationalize observed inertia.[42] Allowing some consumers to be asleep means consumers will stick with the smart default, and cost savings may or may not outweigh acclimation costs. Allowing consumers to consider only a subset of options when awake lowers the acclimation costs necessary to rationalize why a disproportionate share of awake individuals choose the default. Thus, ignoring inattention of either type will cause us to misstate acclimation costs and misstate how sticky smart defaults will be.

The most directly relevant work in the existing literature is Heiss et al. (2016), which also attempts to decompose inertia into inattention- and utility-relevant factors. Our analysis goes beyond this work in three ways. First, we allow for consumers to attend

desirable for unobserved reasons. The second is paperwork costs, where consumers have a cost to choosing any plan that is not the default regardless of whether they have previously chosen that plan.

41. Low-income subsidy beneficiaries have been defaulted into plans with low premiums, although these beneficiaries face substantially less differentiation than regular beneficiaries due to subsidized cost sharing. In Online Appendix C, we use variation from this defaulting among low-income subsidy beneficiaries to aid identification.

42. In Online Appendix C, we examine the robustness of our results to allowing for paperwork costs, that is, costs in arranging to switch plans. This provides a route for a smart default policy to change choice probabilities and cause consumers to stick to the smart default, even in full-consideration models. We find that our welfare conclusions are robust to the existence of paperwork costs as these costs are small relative to acclimation costs.

to only a subset of plans conditional on being awake. Second, we relax assumptions concerning which characteristics are utility-relevant and which are attention-relevant. We show in Online Appendix C.3 that exclusion restrictions relied on in prior work are often rejected.[43] Finally, we consider the implications of our estimates for the (partial-equilibrium) welfare effect of a proposed "smart default" policy.

V.A. Context and Data

Medicare Part D plans provide prescription drug insurance to elderly beneficiaries in the United States. The program was created in 2006 in response to increased spending on pharmaceuticals, creating large out-of-pocket costs for elderly Medicare recipients who at the time had no prescription drug coverage. Our analysis focuses on stand-alone prescription drug insurance plans (PDP plans) and we do not consider plans that provide broader medical insurance (Medicare Advantage). Our main analysis sample consists of 100,000 randomly chosen nondual beneficiaries enrolled in stand-alone plans in 2008–2009. We restrict the sample to beneficiaries for which we observed a prior year plan (dropping 17.5% of beneficiaries). To manage the computational burden of estimating alternative-specific attention parameters, we also restrict our sample only to include plans whose market share is at least 1.5% of plans available in the state in question. This restriction causes us to drop an additional 4% of beneficiary-years, leaving us with a maximum of 17 plans per choice set and 79,286 beneficiaries.

Table II gives the key summary statistics for our sample. The main plan attributes that we observe are annual premiums, deductibles, indicators for whether plans provide coverage in the Part D donut hole, the number of the top 100 drugs included in the formulary, and a quality rating.[44] In addition, we follow

43. In Online Appendix C.3, we explicitly test the alternative restrictions imposed by Heiss et al. (2016) to identify attention probabilities in this setting. These restrictions include assuming that changes in plan attributes do not matter for utility conditional on levels and that demographics such as age affect attention probabilities but not preferences. We statistically reject these assumptions, although we find that models that impose them produce similar attention probabilities (in all cases, cross-derivative asymmetries are contributing to the identification of these probabilities).

44. The donut hole is a gap in coverage included in Part D to lower the fiscal cost of the program. When beneficiaries exceed an initial coverage limit (shifting

TABLE II
SAMPLE DEMOGRAPHICS AND PLAN CHARACTERISTICS

	Mean	Std. dev.
Beneficiary characteristics:		
Age	75.5	8.6
Female	0.643	–
White	0.944	–
Plan characteristics of chosen plans		
Premiums	\$415	\$191
Out-of-pocket costs	\$839	\$713
Deductible	\$59	\$113
Donut hole coverage	0.121	–
Choice characteristics		
Options in Choice set	11.9	2.43
Inertial	93.7	–
No. beneficiary-plan-year	2,261,878	
No. beneficiary observations	79,286	

Notes. This table shows summary statistics for the demographic characteristics and available insurance plans for our sample of Medicare Part D PDP beneficiaries in 2008 and 2009. "Donut hole coverage" refers to either generic or full donut hole coverage.

Abaluck and Gruber (2016) to construct what a beneficiary would pay out of pocket for their claims in all alternative plans in their choice set.[45] The average beneficiary in our sample pays \$839 in annual out-of-pocket costs and \$415 in plan premiums each year. On average, after our sample restrictions, beneficiaries face a choice of 11.9 plans in a given year. Switching between plans is rare in our sample: 93.7% stick with the same plan that they were observed purchasing the previous year (the default plan). This is in line with previous studies of Medicare Part D choice behavior (Heiss, McFadden, and Winter 2010; Abaluck and Gruber 2016). Seven out of 10 Medicare beneficiaries enrolled in these plans during all four annual open enrollment periods from 2006 to 2010

over time between \$2,000 and \$4,000), they must then pay the full cost for the next several thousand dollars in drug costs until reaching the catastrophic coverage threshold (which also varies by year). Some plans offer additional coverage in the donut hole in exchange for higher premiums. The quality rating is based on customer service, member complaints, and "member experience" with the drug plan.

45. To account for uncertainty, we match each individual to 2,000 beneficiaries in the same decile of expenditures in the previous year and then compute the mean and variance of out-of-pocket costs in each plan for all such beneficiaries. More details of this procedure can be found in Abaluck and Gruber (2016).

did not voluntarily switch plans in any of these periods (Hoadley et al. 2013).

1. Reduced-Form Evidence. Before giving our structural estimates, we first provide reduced-form evidence that some inertia is driven by inattention. We treat the plan that an individual chose in the previous year as the default plan. Following Section III, inattention and utility-driven switching costs are separately identified by asymmetries in how the decision to remain inertial depends on default versus rival plan characteristics.[46] To test for such asymmetries, we run a panel regression of an indicator for whether i switched plans in year t on attributes of the default plan and average attributes of alternative plans (with year and beneficiary fixed effects).

$$y_{it} = x_{idt}\alpha_d + (x_{idt} - \bar{x}_{ijt})\alpha_x + \delta_i + \delta_t + e_{it}, \tag{37}$$

where y_{it} is a binary indicator for whether an individual switched from the default at t and $\bar{x}_{ijt}$ is the average of nondefault plans' attributes at t. We consider share-weighted averages (using the choices of new beneficiaries to construct shares), as well as unweighted averages among the three lowest-cost plans, and setting $\bar{x}_{ijt}$ directly equal to attributes of the lowest-cost plan. We report the α_d coefficients, which test whether default attributes are weighted more heavily in switching decisions than rival attributes.

The results are shown in Table III. In all specifications, switching decisions are significantly more sensitive to default premiums and deductibles than to rival attributes (for premiums, the sensitivity to the default is almost three times that of rival attributes). This asymmetry is consistent with the findings of Ho, Hogan, and Scott Morton (2017), who find that consumers do not respond to changes in premiums of the lowest-cost plan, the lowest-cost plan in the same brand, nor the average of the five lowest-cost brands. These patterns of variation are consistent with a model where many consumers do not actively search each period but are induced to make an active choice if the default plan becomes bad enough—in this case, we will see greater responsiveness to attributes of the default plan that affect choices via utility and via prompting consumers actively to

46. See also Moshkin and Shachar (2002).

TABLE III
EXCESS SENSITIVITY TO DEFAULT ATTRIBUTES IN SWITCHING MODEL

	Share weighted	Lowest 3 plans	Lowest-cost plan
Annual premium	0.0801***	0.0931***	0.0914***
(hundreds)	(0.0305)	(0.0348)	(0.0292)
Annual out-of-pocket	0.0057**	– 0.0051	– 0.0096**
costs (hundreds)	(0.0028)	(0.0048)	(0.0046)
Variance of costs	– 0.0075**	0.0040	– 0.0006
(millions)	(0.0037)	(0.0058)	(0.0003)
Deductible (hundreds)	0.1203**	0.1242**	0.1219**
	(0.0545)	(0.0501)	(.0556)
Donut hole coverage	0.0827**	0.0926**	0.0909**
	(0.0410)	(0.0376)	(0.0396)
Average consumer	0.0186	0.0234	0.0210
cost sharing %	(0.0141)	(0.0161)	(0.0143)
Normalized quality	0.0024	0.0479	0.0103
rating	(0.0156)	(0.1663)	(0.0136)
# of top 100 drugs in	0.1669	0.0183	0.1347
formulary	(0.0985)	(0.0157)	(0.1158)

Notes. This table reports coefficients from a panel data regression of an indicator for whether individual i switched at time t on attributes of the default plan, as well as the difference between default plan attributes and rival plan attributes for three different models of rival plan attributes. All models include individual and time fixed effects. In all cases, the reported coefficient is the coefficient on the default attribute conditioning on the difference between the default and rival plan attribute (α_d in equation (37)). The first column computes rival plan attributes as a share-weighted average of nondefault plans where the shares are computed for each (year, state, plan) using the choices of new beneficiaries (who are not included in this regression). The second column uses (unweighted) average attributes among the three lowest-cost plans. The third column uses attributes of the lowest-cost plan for rival plan attributes. The regression also includes an indicator for plans that are missing the *# of top 100 drugs in formulary* variable as well as an interaction of variance of costs and an indicator for individuals with no claims. Standard errors are in parentheses. *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level.

consider other available plans. A notable exception to this pattern is that the coefficients in Table III suggest that consumers are more sensitive to donut hole coverage of rival plans. In our hybrid model, this pattern can be rationalized if consumers are especially likely to attend to attributes of rival plans that offer donut hole coverage.

V.B. Choice Model

To quantify the importance of limited consideration versus utility in rationalizing plan choice, we estimate the hybrid model (Section III.D). Recall that in the hybrid model, consumers are either asleep and choose the default good or, if the default good becomes sufficiently unsuitable, they “wake up” and make an active choice. Conditional on waking up, however, consumers attend only to some of the available options, with the probability of attending to each option depending on the attributes of that option. Following the existing literature, we assume that consumers compare plans only in the current year when making a choice.[47]

As discussed in Section III, the probability of selecting option j, $s_j(\cdot)$ is expressed as:

$$s_0(\mathbf{x}) = \mu(\mathbf{x}_0) \sum_{C \in \mathbb{P}(0)} \prod_{l \in C} \phi_l(\mathbf{x}_l) \prod_{l' \notin C} (1 - \phi_{l'}(\mathbf{x}_{l'})) s_0^\star(\mathbf{x}|C) + (1 - \mu(\mathbf{x}_0))$$

(38)

$$s_j(\mathbf{x}) = \mu(\mathbf{x}_0) \sum_{C \in \mathbb{P}(j)} \prod_{l \in C} \phi_l(\mathbf{x}_l) \prod_{l' \notin C} (1 - \phi_{l'}(\mathbf{x}_{l'})) s_j^\star(\mathbf{x}|C) \quad \text{for } j > 0,$$

where $\mu(\mathbf{x}_0)$ gives the probability of being awake (a function of the attributes of the default good), $\phi_j(\mathbf{x}_j)$ denotes the probability of attending to good j,[48] and $s_j^\star(\mathbf{x}|C)$ denotes the probability of choosing j from choice set C.

There are therefore three sets of parameters to identify and estimate: those that index the probability of waking up, $\mu(\cdot)$; those that index the probability of paying attention to good j conditional on waking up, $\phi_j(\cdot)$; and those that index a consumer's utility function, which determine latent choice shares $s_j^\star(\cdot)$. We assume logistic functional forms for attention probabilties:

$$\phi_j(\mathbf{x}_j) = \frac{\exp(\mathbf{x}_j \gamma)}{1 + \exp(\mathbf{x}_j \gamma)} \tag{39}$$

$$\mu(\mathbf{x}_0) = \frac{\exp(\mathbf{x}_0 \alpha)}{1 + \exp(\mathbf{x}_0 \alpha)}. \tag{40}$$

We assume that the (positive) utility of individual i from choosing plan j at time t is given by:

$$u_{ijt} = \mathbf{x}_{ijt}\beta + \xi \cdot StatusQuo_{ijt} + \epsilon_{ijt}, \tag{41}$$

47. This assumption could be rationalized by assuming that consumers model plans as being static over time, that consumers fail to forecast their inertia, or that consumers are myopic in their plan choices (Dalton, Gowrisankaran, and Town 2020; Abaluck, Gruber, and Swanson 2018 both estimate that Medicare Part D consumers are highly myopic in their drug choices).

48. $\phi_0(\mathbf{x}_0) = 1$ for all $\mathbf{x}_0 \in \chi$.

where ϵ_{ijt} is distributed i.i.d. Type 1 extreme value, $\mathbf{x}_{ijt}\beta$ gives the utility arising from plan characteristics $\mathbf{x}_{ijt}$, and ξ denotes utility-relevant switching costs (acclimation costs) that consumers must pay if they choose any plan other than the status quo from the previous year. Following the earlier literature, we allow consumers to make "errors" by being overly responsive to some plan attributes. Specifically, we allow for separate coefficients on premiums and out-of-pocket costs (although both are in dollar units),[49] and we allow financial plan characteristics to matter for (positive) utility even conditional on their individualized consequences for consumer costs (a rational consumer should only care about deductibles insofar as they affect the distribution of out-of-pocket costs).

V.C. Structural Results

Table IV gives our structural choice model results. We present results for the hybrid model and a conditional logit model that does not allow for inattention. We start by discussing the results for the standard conditional logit model (Table IV, column (1)). The stylized facts from (Abaluck and Gruber 2011, 2016) are apparent. Consumers weigh premiums more heavily than out-of-pocket costs, and consumers appear responsive to plan attributes such as deductibles even after we control for the financial impact of those deductibles via out-of-pocket costs. The fact that consumers are overwhelmingly likely to choose the default plan implies acclimation costs of $1,224 under the conditional logit model.

Next we estimate the hybrid model on the same sample. After adjusting for limited attention, this model implies acclimation costs of $287; less than a quarter of the size of those estimated assuming full consideration. In the conditional logit model, we estimate that consumers are risk-loving. However, allowing for limited consideration, they appear risk-averse. We also generally find that consumers are more responsive to plan attributes conditional on consideration than is implied by the conditional logit model.

Our consideration coefficients imply that on average, consumers consider only 1.97 available plans each year. This is driven both by consumers not actively searching each period and thus

49. Appendix C of Abaluck and Gruber (2009) shows how this pattern can be derived from a model where some consumers are imperfectly informed about out-of-pocket costs.

TABLE IV
CONDITIONAL LOGIT AND HYBRID MODEL

	Conditional logit (1)	Hybrid model: Utility (2)	Hybrid model: $\phi(\cdot)$ (3)	Hybrid model: $\mu(\cdot)$ (4)
Annual premium (hundreds)	-0.4298^{***}	-0.6293^{***}	-1.118^{***}	0.0281^{**}
	(0.0103)	(0.0331)	(0.0731)	(0.0116)
Annual out-of-pocket costs (hundreds)	-0.1420^{***}	-0.0054	-0.3797^{***}	0.0657^{***}
	(0.0131)	(0.0321)	(0.0697)	(0.0206)
Variance of costs (millions)	0.1626^{***}	-0.1687^{***}	0.5501^{***}	1.5589^{***}
	(0.0183)	(0.0492)	(0.0981)	(0.0301)
Deductible (hundreds)	-0.4108^{***}	-0.6642^{***}	-0.0531	0.2008^{***}
	(0.021)	(0.0552)	(0.1017)	(0.0483)
Donut hole coverage	0.5656^{***}	1.3579^{***}	10.1685^{***}	-0.0889
	(0.0609)	(0.2071)	(0.7432)	(0.1127)
Average consumer cost sharing %	-0.2209^{***}	-1.2484^{***}	1.3849^{***}	-0.3911^{***}
	(0.0242)	(0.0678)	(0.1474)	(0.0663)
# of top 100 drugs in formulary	0.1243^{***}	0.0010	0.4557^{***}	-0.1157^{***}
	(0.0122)	(0.0362)	(0.0483)	(0.0371)
Normalized quality rating	0.0399^{***}	-0.0066	0.1976^{***}	-0.0843^{***}
	(0.0021)	(0.0062)	(0.0181)	(0.0043)
Default	5.2599^{***}	1.8052^{***}		
	(0.0224)	(0.1473)		
Constant			1.626^{***}	-1.853^{***}
			(0.2126)	(0.0764)
Switching cost	\$1,224	\$287		
Mean attention probability			67.46%	15.09%

Notes. This table gives estimates from the conditional logit and "hybrid" version of the consideration set framework. Estimates in all models are the coefficients in the utility and attention equations (not marginal effects). The coefficients in the DSC component of the model, $\mu(\cdot)$, are the coefficients on the listed characteristics of the default good. The coefficients in the ASC component of the model, $\phi(\cdot)$, are the coefficients on the listed characteristics of good j on the likelihood of j being considered. The model also includes an indicator for plans that are missing the # *of top 100 drugs in formulary* variable as well as an interaction of variance of costs and an indicator for individuals with no claims. Standard errors are in parentheses. *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level.

choosing their default option with probability 1 (μ) and by them only considering a subset of available plans if they do search (ϕ). Consider first the estimates of the effect of plan attributes on μ, the probability of being awake. These estimates imply that the majority of consumers do not actively search each period: only 15% of consumers consider a plan other than the default option. The effects of default characteristics on the probability of waking up are intuitive. Consumers are more likely to wake up and consider alternative plans when premiums, out-of-pocket costs, cost variance, or deductibles increase and when there is a decrease in quality ratings or the number of top 100 drugs which are covered. Conditional on waking up, we find that consumers attend to 7.16 plans in their choice set on average. We find that consumers are more likely to attend to options with lower premiums and out-of-pocket costs, and those with higher quality ratings, more of the top 100 drugs on their formularies, and greater donut hole coverage. The sensitivity of consideration probabilities to donut hole coverage is especially notable—if a rival plan has donut hole coverage, consumers are about 10 times more likely to consider it.

V.D. Welfare Analysis of Smart Defaults

Motivated partly by the considerable inertia in insurance plan choice patterns described above, Handel and Kolstad (2015) propose a smart default policy in which an individual is switched into the lowest-cost plan available in each year provided that their monetary gain from such a switch exceeds some threshold.[50] Under this proposal, all enrollees would retain the ability to opt out of their new default and either switch back to their original plan or choose any of the alternative plans available. This policy has not been implemented. Therefore, we must rely on existing variation in the data and structural methods to predict how consumers would respond to the policy and normatively evaluate the results.

1. Normative Utility. To evaluate the smart default policy, we must take a stand on what is relevant for normative utility. Following Abaluck and Gruber (2011) and Heiss, McFadden, and Winter (2010),[51] we assume that, apart from switching costs (discussed below), normative utility depends only on total cost, risk protection, and observable quality measures. We denote this utility by v_{ij} (we suppress the subscript t, although plan attributes vary over time as well).[52] In other words, normative utility is given by (the negative) of expected out-of-pocket costs, plus the

50. In theory, we could also consider a policy where consumers are only switched if their utility gain exceeds some threshold. Our key results go through under this modification, although the monetary proposal has more practical relevance due to its transparency.

51. This specification is also supported by our finding of no spurious state dependence in Online Appendix C.

52. Formally, $v_{ij} = \pi_j + \mu^*_{ij} + \frac{\beta_2}{\beta_0}\sigma^2_{ij} + \frac{\delta}{\beta_0}q_j$.

dollar-equivalent risk protection and the dollar-equivalent plan quality rating (where, in each case, the dollar-equivalent measures are computed by normalizing by the coefficient on premiums).

In our baseline results, we assume that utility-relevant switching costs are all acclimation costs. In other words, these are costs that must be paid when a beneficiary enrolls in a plan with which they do not previously have experience.[53] In Online Appendix C, we consider several other possible drivers of inertia and discuss in more detail how our analysis fits with alternative attempts to decompose inertia into different mechanisms. One alternative story is that inertia may be driven by the paperwork costs required to fill out and send in the forms required to switch plans (Luco 2019). Paperwork costs have different counterfactual implications than acclimation costs: paperwork costs make it costly to return to the original plan after being defaulted away, whereas acclimation costs make it costly to remain in a new plan even if it otherwise saves money. With positive paperwork costs, a smart default policy can affect choice behavior in full-consideration models as it will now be costly to opt out of the policy. Another possibility is that inertia arises from "spurious state dependence," that is, something unobservably good about chosen plans. To separately identify these factors, our analysis in the Online Appendix exploits additional variation from the random assignment of a subset of beneficiaries (low-income subsidy beneficiaries) to alternative plans. This analysis suggests that both paperwork costs and spurious state dependence (at the brand level) play a limited role in explaining observed inertia.[54]

2. Welfare Change. Let consumer i's original plan, or old default, be denoted by the subscript o. The expected welfare change,

53. This is a form of what Heckman (1981) calls structural state dependence, wherein choices directly affect choice probabilities.

54. Specifically, when low-income subsidy recipients no longer qualify for low-income subsidies, they must pay paperwork costs but not acclimation costs to return to their original plans. We find that they nonetheless do so at a disproportionate rate, suggesting that paperwork costs are small. In addition, we find that these beneficiaries are not more likely to return to other plans from the same brand, suggesting that persistent unobserved heterogeneity at the brand level is small as well.

$\triangle W_i$, can be expressed as:

$$\begin{aligned} \triangle W_i &= W_i^1 - W_i^0 \\ &= \xi \triangle s_{io} + \sum_j \triangle s_{ij} v_{ij}, \end{aligned} \tag{42}$$

where ξ is the acclimation cost and $\triangle s_{ij} = s_{ij}^1 - s_{ij}^0$, where superscripts represent either the current scenario (0) or the counterfactual policy scenario (1).

Defaults change welfare through two channels. First, acclimation costs will be paid by anyone who switches away from the original plan as a result of the new default (we will generally have $\triangle s_{io} < 0$). Second, the policy will change choice probabilities given its aim of inducing people to choose lower-cost plans (which might have a higher v_{ij}). Estimating consideration probabilities is required to bring equation (42) to the data for two reasons. First, estimates of ξ will depend on the degree of inattention. Second, to recover Δs_{ij}, we need to simulate the effect of smart defaults on choice probabilities; this will depend on the degree of inattention as well as the structural preference parameters. Inattention will tend to make smart defaults stickier in the sense that consumers will not return to their original plans.

The above allows for the smart default policy to change consideration probabilities given that the new default will have different characteristics and therefore a different $\mu(\cdot)$. However, it does not allow for the smart default policy to have a direct effect on attention. Defaulting consumers to a different plan might directly wake them up (especially given any outreach campaign that might realistically accompany such a policy). Alternatively, if consumers are rationally inattentive, they may be less likely to pay attention if the smart default is even more suitable (Carroll et al. 2009). This matters for positive and normative reasons. Positively, if smart defaults cause people to wake up, we may see more people revisiting the original plan or making an active choice than we would otherwise predict. Normatively, if smart defaults shift the degree of inattention, we might worry that we are imposing an additional effort cost on some consumers. This additional effort cost is not identified in our model without further assumptions about what drives the decision to pay attention (e.g., it could be identified if we imposed rational inattention). In our results, we thus evaluate smart defaults under a range of assumptions about how attention

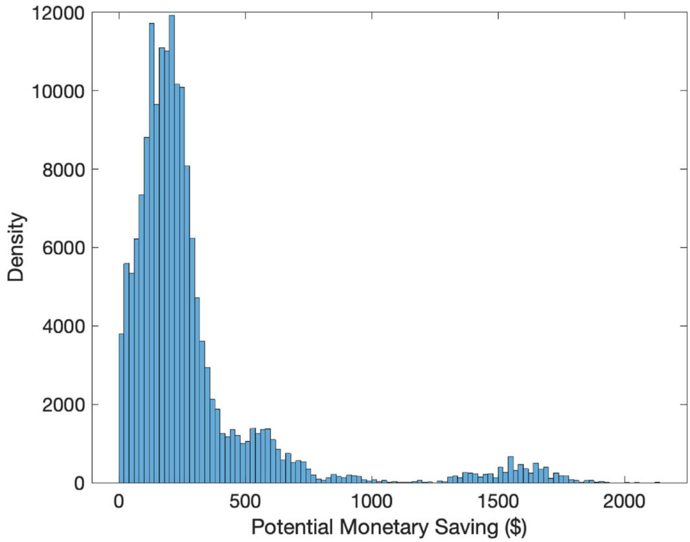

FIGURE III

Distribution of Savings

This figure shows the distribution of potential monetary savings (annual premiums plus estimated out-of-pocket costs) from being switched from a recipient's current plan to the lowest-cost plan.

is directly perturbed and under a range of values for the effort cost of paying attention.[55]

V.E. Simulation Results

Figure III gives the distribution of cost savings achieved under the policy. Note that this ignores the effect of switching costs or other utility-relevant factors. On average, beneficiaries save \$286 from switching to the lowest-cost plan in their choice set. However, there are some consumers with substantially greater potential savings; for example, 5% of beneficiaries could save more than \$1,000.

Looking at cost savings alone ignores the effect of acclimation costs associated with switching consumers that lower the utility

55. This approach is similar to Goldin and Reck (2020), who advocate assuming normative switching costs are a fraction of positive switching costs between 0 and 1. We instead attempt to identify normative switching costs other than effort costs, and consider the robustness of our model to different assumptions about these effort costs.

TABLE V
WELFARE IMPACT OF A SMART DEFAULT POLICY

	Market shares conditional on being switched			Attention cost			
	Smart default	Previous plan	Other plan	$0	$50	$100	$200
Conditional logit	1.000	0.000	0.000	$0	$0	$0	$0
Hybrid model	0.961	0.001	0.048	$46	$42	$37	$29
Direct effect on μ							
25%	0.946	0.002	0.053	$39	$32	$25	$12
50%	0.902	0.003	0.095	$18	$2	−$14	−$46
75%	0.853	0.005	0.142	−$3	−$30	−$57	−$111
100%	0.804	0.006	0.189	−$25	−$63	−$101	−$177
Proportion Switched by Policy: 89%							

Notes. This table shows overall welfare effects of a smart default policy where consumers are switched to the lowest-cost plan available in their market. Each row shows results with different assumptions about the direct effect of the smart default policy on the probability of paying attention, and each column shows alternative assumptions about the cost of paying attention.

benefits from moving plans. Furthermore, total cost is not the only relevant characteristic for normative utility (i.e., risk aversion and quality ratings also matter). We simulate welfare under the smart default policy conditional on the structural parameters reported in Table IV. To simulate smart defaults, we switch a consumer to the lowest-cost plan available but then allow them to either switch back to their original plan or some other plan. If they choose a plan other than their original plan, they must pay acclimation costs. Consumers might be made worse off by this policy if inattentive consumers are inadvertently driven to pay acclimation costs that outweigh the gains from being enrolled in an otherwise better plan.

Table V gives our baseline results. Only 11% of consumers are in the lowest-cost plan and thus 89% of beneficiaries are switched by the policy to the lowest-cost plan, and among these, 96% stick with the new default. When we use the acclimation costs estimated in the hybrid model, we find this policy has small but positive effects on welfare. This reflects the fact that the utility benefits of the lowest-cost plan on average outweigh our estimated switching costs.[56] Note that in the full-consideration conditional logit model, this policy would have no effect on choice behavior or welfare. Although acclimation costs are large in this model

56. The lowest-cost plan also often has better noncost characteristics than a consumer's original plan.

($1,224), a full-consideration model predicts that all consumers would switch back to their original plan as, absent inattention, defaults have no effect on behavior in our model.

These results still ignore the possibility that smart defaults might induce people to pay attention, which may itself be costly. Although this cost is not identified in our model without additional structure, we consider alternative assumptions about the induced attention probability μ and the cost of paying attention. As the direct effect of the policy on the likelihood of waking up increases, defaulted consumers become more likely to wake up, pay attention costs, and potentially switch to a new plan. While this could in theory lead to a positive effect on welfare, we find that consumers tend to switch to worse plans on average from the perspective of normative utility.

An alternative policy would only reassign those consumers who stand to benefit the most from reassignment. What if we reassign only those beneficiaries for whom the potential cost savings exceed our estimated acclimation costs? The results are shown in Table VI. In this case, the average benefits almost triple and remain positive for any plausible assumptions about attention costs for consumers who are switched by the policy. We now reassign far fewer beneficiaries, but the benefits conditional on being reassigned increase substantially to an average of $400–$500 per reassigned beneficiary.

One important caveat to these results is that we consider only a partial-equilibrium analysis: premiums and plan attributes are held constant. In practice, defaulting a large number of beneficiaries into alternative plans would likely cause firms to respond by altering their premiums and coverage characteristics. Decarolis (2015) highlights one way such incentives can backfire in a context where the government pays premiums. In the more general context, Ho, Hogan, and Scott Morton (2017) suggest that reducing inertia should enhance competition between plans and lower premiums. Nonetheless, the responsiveness of plan attributes to changes in inertia is not fully understood.

If firms lowered the cost of their plans so that beneficiaries would be defaulted into their product while simultaneously reducing plan desirability on other dimensions such as plan quality, then the welfare benefits of the policy might be diminished. To combat this, one might consider a smart default policy in which beneficiaries are only assigned to plans that also have high

TABLE VI

WELFARE EFFECT OF A SMART DEFAULT POLICY: RESTRICTED REASSIGNMENT TO THOSE WITH SAVINGS GREATER THAN SWITCHING COSTS

	Market shares conditional on being switched			Attention cost			
	Smart default	Previous plan	Other plan	$0	$50	$100	$200
Mean welfare change: full sample							
Hybrid parameters				$114	$113	$112	$111
Direct effect on attention probability							
25%				$109	$107	$106	$102
50%				$94	$90	$85	$76
75%				$79	$71	$64	$49
100%				$63	$52	$42	$21
Proportion consumers switched by policy: 24%							
Mean welfare change: conditional on being switched							
Hybrid parameters	0.947	0.001	0.052	$479	$476	$472	$465
Direct effect on attention probability							
25%	0.923	0.002	0.076	$458	$451	$444	$429
50%	0.852	0.003	0.145	$395	$377	$358	$320
75%	0.778	0.005	0.217	$330	$299	$267	$205
100%	0.705	0.006	0.289	$264	$220	$177	$89

Notes. This table shows overall welfare effects of a smart default policy where consumers who will experience a monetary saving of at least ξ are switched to the lowest-cost plan available in their market. Each row shows results with different assumptions about the direct effect of the smart default policy on the probability of paying attention, while each column shows alternative assumptions about the cost of paying attention.

quality ratings (based on beneficiary feedback). In Online Appendix Table 11, we consider such an alternative policy. We again only reassign beneficiaries whose potential cost savings exceed acclimation costs, this time reassigning beneficiaries to the lowest-cost plan with a quality rating in the top quartile of available plans. In this case, the welfare benefits are actually slightly smaller than in our baseline simulation (because these new plans are higher cost).

Our analyses reported here do not allow for the possibility that in addition to acclimation costs, consumers face paperwork costs that prevent them from switching plans even conditional on paying attention. When paperwork costs are high, even consumers who pay attention may become stuck in unsuitable plans because they do not want to bother to switch. In Online Appendix C, we consider a generalization of the model here in which consumers face both paperwork and acclimation costs, which we identify using the random assignment of low-income subsidy enrollees to alternative plans (focusing specifically on their choices when they are no longer eligible for such subsidies). The upshot of that analysis is that paperwork costs are negligible, so our conclusions here are unchanged.

VI. Conclusion

Discrete-choice models with consideration sets relax the strong assumption that consumers consider all of the options available to them before making a choice. In the applied literature to date, such models have been identified either by bringing in auxiliary information on what options consumers consider or assuming that some characteristics impact attention or utility but not both. This article shows that these assumptions are not required for identification. We show that a broad class of such models are identified from variation already available in the data. Consideration set probabilities are constructively identified by deviations from Slustky symmetry, that is, asymmetries in the matrix of cross-derivatives of choice probabilities with respect to characteristics of rival goods.

To highlight the power of this result, we use our framework to model limited consideration in Medicare Part D, a setting with considerable inertia in choice behavior. Our model allows for two forms of inattention: we allow for consumers to be asleep and simply to choose their default option, and awake to only consider a subset of the available plans. Our results show that, while most inertia is driven by inattention in Part D, there remain nontrivial utility-relevant adjustment costs. We simulate the welfare effect of a "smart default" policy, finding that defaulting consumers into lower-cost plans can produce large benefits. This is in contrast to models that assume full consideration, which predict that a smart default policy would have no effect on choice behavior and consumer welfare.

While we show that deviations from Slutsky symmetry are indicative of imperfect attention in a large class of models, our constructive identification results use the additional structure imposed by the widely applied DSC and ASC frameworks. One direction for future work is to characterize more generally when consideration probabilities can be recovered from choice data. One important case is the K-rank models considered in Honka (2014) and Honka, Hortaçsu, and Vitorino (2017) in which consumers consider the K-goods which are highest ranked according to some index, thus violating the independence assumption of the ASC model. In addition, while we consider consideration at the level of goods, an important question for future work is to characterize the conditions under which choice data suffices to recover inattention at the level of attributes (as in Kőszegi and Szeidl 2012; Bordalo, Gennaioli, and Shleifer 2013).[57] We hope that the sufficient conditions given here will make it possible to adapt consideration set models to a wider range of settings than they have previously been applied to.

The model we consider provides a general empirical framework for analyzing policies which change defaults across a wide variety of settings such as 401(k) plans, insurance, school choice, and online consumer purchases, among others (Carroll et al. 2009; Bernheim, Fradkin, and Popov 2015). Our model implies that defaults change behavior because of inattention. By allowing awake individuals to consider only a subset of options, we avoid overstating utility-relevant switching costs. Rather than imposing specific models of rational inattention, our model estimates consideration probabilities from the data. Without an explicit microfoundation for attention, we cannot predict how attention might shift as we change the underlying context, but we show in our application that our welfare conclusions are robust to alternative assumptions about such shifts.

57. Abaluck and Compiani (2020) produces results along these lines.

With additional structure, consideration set models can be used to identify parameters of interest such as search costs, and they enable us to construct counterfactuals and explore normative questions that would not be possible in conventional models. We can ask, for example, how might beneficiaries choose if they considered all available options? When choices correlate with cognitive ability, is this because cognitive ability affects preferences or because it affects consumers' ability to consider all options? Do some demographic or choice set features increase the likelihood that consumers are attentive? How much better off might consumers be if we switched them to alternative options? We hope that future work will explore these questions in more detail in other policy-relevant scenarios.

YALE UNIVERSITY, UNITED STATES
UNIVERSITY OF OXFORD, UNITED KINGDOM

SUPPLEMENTARY MATERIAL

An Online Appendix for this article can be found at the *Quarterly Journal of Economics* online.

DATA AVAILABILITY

Data and code replicating the figures and tables in this article can be found in Abaluck and Adams-Prassl (2021) in the Harvard Dataverse, https://doi.org/10.7910/DVN/ZVYOCQ.

REFERENCES

Abaluck, Jason, and Abi Adams-Prassl, "Replication Data for: 'What Do Consumers Consider before They Choose? Identification from Asymmetric Demand Responses'," (2021), Harvard Dataverse, https://doi.org/10.7910/DVN/ZVYOCQ.

Abaluck, Jason, and Giovanni Compiani, "A Method to Estimate Discrete Choice Models that is Robust to Consumer Search," NBER Technical Report, 2020.

Abaluck, Jason T., and Jonathan Gruber, "Choice Inconsistencies Among the Elderly: Evidence from Plan Choice in the Medicare Part D Program," NBER Working Paper no. 14759, 2009.

———, "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program," *American Economic Review*, 101 (2011), 1180–1210.

———, "Evolving Choice Inconsistencies in Choice of Prescription Drug Insurance," *American Economic Review*, 106 (2016), 2145–2184.

Abaluck, Jason, Jonathan Gruber, and Ashley Swanson, "Prescription Drug Use under Medicare Part D: A Linear Model of Nonlinear Budget Sets," *Journal of Public Economics*, 164 (2018), 106–138.

Aguiar, Victor H., Maria Boccardi, Nail Kashaev, and Jeongbin Kim, “Does Random Consideration Explain Behavior When Choice Is Hard? Evidence from a Large-Scale Experiment,” University of Western Ontario Working Paper, 2018.

Aguiar, Victor H., and Roberto Serrano, “Slutsky Matrix Norms: The Size, Classification, and Comparative Statics of Bounded Rationality,” *Journal of Economic Theory*, 172 (2017), 163–201.

Anderson, Simon P., Andre De Palma, and Jacques-Francois Thisse, *Discrete Choice Theory of Product Differentiation* (Cambridge, MA: MIT Press, 1992).

Barseghyan, Levon, Maura Coughlin, Francesca Molinari, and Joshua Teitelbaum, “Heterogeneous Consideration Sets and Preferences,” arVix preprint 1907.02337, 2021.

Barseghyan, Levon, Francesca Molinari, and Matthew Thirkettle, “Discrete Choice under Risk with Limited Consideration,” arVix preprint 1902.06629, 2021.

Ben-Akiva, Moshe, and Bruno Boccara, “Discrete Choice Models with Latent Choice Sets,” *International Journal of Research in Marketing*, 12 (1995), 9–24.

Bernheim, B. Douglas, Andrey Fradkin, and Igor Popov, “The Welfare Economics of Default Options in 401(k) Plans,” *American Economic Review*, 105 (2015), 2798–2837.

Berry, Steven T., and Philip A. Haile, “Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers,” NBER Working Paper 15276, 2009.

———, “Identification in Differentiated Products Markets Using Market Level Data,” *Econometrica*, 82 (2014), 1749–1797.

———, “Identification in Differentiated Products Markets,” *Annual Review of Economics*, 8 (2016), 27–52.

Berry, Steven, James Levinsohn, and Ariel Pakes, “Automobile Prices in Market Equilibrium,” *Econometrica*, 63 (1995), 841–890.

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor, “Choose to Lose: Health Plan Choices from a Menu with Dominated Options,” *Quarterly Journal of Economics*, 132 (2017), 1319–1372.

Bierlaire, Michel, Denis Bolduc, and Daniel McFadden, “The Estimation of Generalized Extreme Value Models from Choice-Based Samples,” *Transportation Research Part B: Methodological*, 42 (2008), 381–394.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, “Salience and Consumer Choice,” *Journal of Political Economy*, 121 (2013), 803–843.

Bruins, Marianne, James A. Duffy, Michael P Keane, and Anthony A Smith, Jr., “Generalized Indirect Inference for Discrete Choice Models,” *Journal of Econometrics*, 205 (2018), 177–203.

Caplin, Andrew, Mark Dean, and John Leahy, “Rational Inattention, Optimal Consideration Sets and Stochastic Choice,” *The Review of Economic Studies*, 86 (2019), 1061–1094.

Carroll, Gabriel D., James J. Choi, David Laibson, Brigitte C. Madrian, and Andrew Metrick, “Optimal Defaults and Active Decisions,” *Quarterly Journal of Economics*, 124 (2009), 1639–1674.

Cattaneo, Matias D., Xinwei Ma, Yusufcan Masatlioglu, and Elchin Suleymanov, “A Random Attention Model,” *Journal of Political Economy*, 128 (2020), 2796–2836.

Compiani, Giovanni, and Yuichi Kitamura, “Using Mixtures in Econometric Models: A Brief Review and Some New Results,” *Econometrics Journal*, 19 (2016), C95–C127.

Conlon, Christopher T., and Julie Holland Mortimer, “Demand Estimation under Incomplete Product Availability,” *American Economic Journal: Microeconomics*, 5 (2013), 1–30.

Crawford, Gregory S., Rachel Griffith, and Alessandro Iaria, “A Survey of Preference Estimation with Unobserved Choice Set Heterogeneity,” *Journal of Econometrics*, 222 (2021), 4–43.

Dalton, Christina M., Gautam Gowrisankaran, and Robert Town, “Myopia and Complex Dynamic Incentives: Evidence from Medicare Part D,” *Review of Economic Studies*, 87 (2020), 822–869.

Daly, Andrew, and Stanley Zachary, “Improved Multiple Choice Models,” *Determinants of Travel Choice*, 335 (1978), 357.
Dardanoni, Valentino, Paola Manzini, Marco Mariotti, and Christopher J. Tyson, “Inferring Cognitive Heterogeneity from Aggregate Choices,” *Econometrica*, 88 (2020), 1269–1296.
Davis, Peter, and Pasquale Schiraldi, “The Flexible Coefficient Multinomial Logit (FC-MNL) Model of Demand for Differentiated Products,” *RAND Journal of Economics*, 45 (2014), 32–63.
Decarolis, Francesco, “Medicare Part D: Are Insurers Gaming the Low Income Subsidy Design?,” *American Economic Review*, 105 (2015), 1547–1580.
Draganska, Michaela, and Daniel Klapper, “Choice Set Heterogeneity and the Role of Advertising: An Analysis with Micro and Macro Data,” *Journal of Marketing Research*, 48 (2011), 653–669.
Fox, Jeremy T., David H. Hsu, and Chenyu Yang, “Unobserved Heterogeneity in Matching Games with an Application to Venture Capital,” NBER Technical Report, 2012.
Gabaix, Xavier, “A Sparsity-Based Model of Bounded Rationality,” *Quarterly Journal of Economics*, 129 (2014), 1661–1710.
———, “Behavioral Inattention,” in *Handbook of Behavioral Economics, Vol. 2*, B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds. (Elsevier, 2019), 261–343.
Gaynor, Martin, Carol Propper, and Stephan Seiler, “Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service,” *American Economic Review*, 106 (2016), 3521–3557.
Goeree, Michelle Sovinsky, “Limited Information and Advertising in the US Personal Computer Industry,” *Econometrica*, 76 (2008), 1017–1074.
Goldin, Jacob, and Daniel Reck, “Optimal Defaults with Normative Ambiguity,” *Review of Economics and Statistics*, 2020, 1–45.
Gourieroux, Christian, Alain Monfort, and Eric Renault, “Indirect Inference,” *Journal of Applied Econometrics*, 8 (1993), S85–S118.
Handel, Benjamin R., “Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts,” *American Economic Review*, 103 (2013), 2643–2682.
Handel, Ben, and Jonathan Kolstad, “Getting the Most from Marketplaces: Smart Policies on Health Insurance Choices,” *Hamilton Project Discussion Paper 2015-08*, 2015.
Hastings, Justine S., and Jesse M. Shapiro, “Fungibility and Consumer Choice: Evidence from Commodity Price Shocks,” *Quarterly Journal of Economics*, 128 (2013), 1449–1498.
Hauser, John R., and Birger Wernerfelt, “An Evaluation Cost Model of Consideration Sets,” *Journal of Consumer Research*, 16 (1990), 393–408.
Heckman, James J., “Heterogeneity and State Dependence,” in *Studies in Labor Markets,* Sherwin Rosen, ed. (Chicago: University of Chicago Press, 1981), 91–140.
Heckman, James J., and Edward Vytlacil, “Structural Equations, Treatment Effects, and Econometric Policy Evaluation,” *Econometrica*, 73 (2005), 669–738.
Heiss, Florian, Adam Leive, Daniel McFadden, and Joachim Winter, “Plan Selection in Medicare Part D: Evidence from Administrative Data,” *Journal of Health Economics*, 32 (2013), 1325–1344.
Heiss, Florian, Daniel McFadden, and Joachim Winter, “Mind the Gap! Consumer Perceptions and Choices of Medicare Part D Prescription Drug Plans,” in David A. Wise, ed. *Research Findings in the Economics of Aging,* (Chicago: University of Chicago Press, 2010, 413–481.
Heiss, Florian, Daniel McFadden, Joachim Winter, Amelie Wupperman, and Bo Zhou, “Inattention and Switching Costs as Sources of Inertia in Medicare Part D,” NBER Working Paper 22765, 2016.
Henry, Marc, Yuichi Kitamura, and Bernard Salanié, “Partial Identification of Finite Mixtures in Econometric Models,” *Quantitative Economics*, 5 (2014), 123–144.

Ho, Kate, Joseph Hogan, and Fiona Scott Morton, “The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program,” *RAND Journal of Economics*, 48 (2017), 877–905.

Hoadley, Jack, Elizabeth Hargrave, Laura Summer, Juliette Cubanski, and Tricia Neuman, “To Switch or not to Switch: Are Medicare Beneficiaries Switching Drug Plans to Save Money?,” Kaiser Family Foundation Working Paper, 2013.

Honka, Elisabeth, “Quantifying Search and Switching Costs in the US Auto Insurance Industry,” *RAND Journal of Economics*, 45 (2014), 847–884.

Honka, Elisabeth, and Pradeep Chintagunta, “Simultaneous or Sequential? Search Strategies in the us Auto Insurance Industry,” *Marketing Science*, 36 (2016), 21–42.

Honka, Elisabeth, Ali Hortaçsu, and Maria Ana Vitorino, “Advertising, Consumer Awareness, and Choice: Evidence from the US Banking Industry,” *RAND Journal of Economics*, 48 (2017), 611–646.

Hortaçsu, Ali, Seyed Ali Madanizadeh, and Steven L. Puller, “Power to Choose? An Analysis of Consumer Inertia in the Residential Electricity Market,” *American Economic Journal: Economic Policy*, 9 (2017), 192–226.

Kawaguchi, Kohei, Kosuke Uetake, and Yasutora Watanabe, “Designing Context-Based Marketing: Product Recommendations under Time Pressure,” *Management Science*, forthcoming.

Keane, Michael, and Anthony A. Smith, “Generalized Indirect Inference for Discrete Choice Models,” Yale University Working Paper, 2003.

Koning, Ruud H., and Geert Ridder, “Discrete Choice and Stochastic Utility Maximization,” *Econometrics Journal*, 6 (2003), 1–27.

Kőszegi, Botond, and Adam Szeidl, “A Model of Focusing in Economic Choice,” *Quarterly Journal of Economics*, 128 (2012), 53–104.

Lewbel, Arthur, “Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables,” *Journal of Econometrics*, 97 (2000), 145–177.

———, “Endogenous Selection or Treatment Model Estimation,” *Journal of Econometrics*, 141 (2007), 777–806.

Lewbel, Arthur, and Xun Tang, “Identification and Estimation of Games with Incomplete Information Using Excluded Regressors,” *Journal of Econometrics*, 189 (2015), 229–244.

Lu, Zhentong, “Estimating Multinomial Choice Models with Unobserved Choice Sets,” Unpublished Manuscript, 2016.

Luco, Fernando, “Switching Costs and Competition in Retirement Investment,” *American Economic Journal: Microeconomics*, 11 (2019), 26–54.

Magnac, Thierry, and Eric Maurin, “Identification and Information in Monotone Binary Models,” *Journal of Econometrics*, 139 (2007), 76–104.

Manski, Charles F., “The Structure of Random Utility Models,” *Theory and Decision*, 8 (1977), 229–254.

Manski, Charles F., and Steven R. Lerman, “The Estimation of Choice Probabilities from Choice Based Samples,” *Econometrica: Journal of the Econometric Society*, 45 (1977), 1977–1988.

Manzini, Paola, and Marco Mariotti, “Stochastic Choice and Consideration Sets,” *Econometrica*, 82 (2014), 1153–1176.

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y. Ozbay, “Revealed Attention,” *American Economic Review*, 102 (2012), 2183–2205.

Matejka, Filip, and Alisdair McKay, “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model,” *American Economic Review*, 105 (2014), 272–298.

Matzkin, Rosa L., “Nonparametric Identification,” *Handbook of Econometrics*, 6 (2007), 5307–5368.

Moshkin, Nickolay V., and Ron Shachar, “The Asymmetric Information Model of State Dependence,” *Marketing Science*, 21 (2002), 435–454.

Reutskaja, Elena, Rosemarie Nagel, Colin F. Camerer, and Antonio Rangel, “Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study,” *American Economic Review*, 101 (2011), 900–926.

Shocker, Allan D., Moshe Ben-Akiva, Bruno Boccara, and Prakash Nedungadi, "Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models, and Suggestions," *Marketing Letters*, 2 (1991), 181–197.

Smith, Anthony A., "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8 (1993), S63–S84.

Swait, Joffre, and Moshe Ben-Akiva, "Incorporating Random Constraints in Discrete Models of Choice Set Generation," *Transportation Research Part B: Methodological*, 21 (1987), 91–102.

Tamer, Elie, "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70 (2003), 147–165.

Treisman, Anne M., and Garry Gelade, "A Feature-Integration Theory of Attention," *Cognitive Psychology*, 12 (1980), 97–136.

Van Nierop, Erjen, Bart Bronnenberg, Richard Paap, Michel Wedel, and Philip Hans Franses, "Retrieving Unobserved Consideration Sets from Household Panel Data," *Journal of Marketing Research*, 47 (2010), 63–74.

CALL FOR NOMINATIONS

$300,000 Nemmers Prize in Economics

Northwestern University invites nominations for the Erwin Plein Nemmers Prize in Economics, to be awarded during the 2026–27 academic year. The prize pays the recipient $300,000. Recipients of the Nemmers Prize present lectures, participate in department seminars, and engage with Northwestern faculty and students in other scholarly activities.

Details about the prize and the nomination process can be found at **nemmers.northwestern.edu**. Candidacy for the Nemmers Prize is open to those with careers of outstanding achievement in their disciplines as demonstrated by major contributions to new knowledge or the development of significant new modes of analysis.Individuals of all nationalities and institutional affiliations are eligible except current or recent members of the Northwestern University faculty and past recipients of the Nemmers or Nobel Prize. **Nominations will be accepted until January 14, 2026.**

The Nemmers prizes are made possible by a generous gift to Northwestern University by the late Erwin Esser Nemmers and the late Frederic Esser Nemmers.

Northwestern

Nemmers Prizes • Office of the Provost • Northwestern University • Evanston, Illinois 60208
nemmers.northwestern.edu