Journal of Behavioral Decision Making

WILEY

RESEARCH ARTICLE OPEN ACCESS

Choice Modeling With Context Effects: Generalization for Observational Data

Zakaria Babutsidze[1] | William Rand[2] | Emil Mirzayev[3] | Nobuyuki Hanaki[4,5] | Ismael Rafaï[6] | Thierry Delahaye[7] | Rodrigo Acuna-Agost[8]

[1]SKEMA Business School, Université Côte d'Azur (GREDEG), France | [2]North Carolina State University, Raleigh, North Carolina, USA | [3]University College London, London, UK | [4]University of Osaka, Osaka, Japan | [5]University of Limassol, Limassol, Cyprus | [6]CNRS, Université Toulouse Capitole, Toulouse School of Management and Toulouse School of Economics, Toulouse, France | [7]Amadeus IT Pacific, Sydney, Australia | [8]Amadeus SAS, Sophia Antipolis, France

Correspondence: Zakaria Babutsidze (zakaria.babutsidze@skema.edu)

Received: 5 June 2024 | **Revised:** 30 June 2025 | **Accepted:** 30 June 2025

Funding: This study was funded by the Agence Nationale de la Recherche, the French National Research Agency (grant numbers ANR-15-IDEX-01 and ANR-17-EURE-004).

Keywords: behavioral decision making | choice modeling | clustering | context effects | observational data

ABSTRACT

Established procedures of analyzing the effect of context on choice consider simple, compact environments in laboratory settings. However, these approaches severely limit the study of context effects and, as a consequence, the applicability of their findings. In this paper, we generalize existing approaches in modeling choice with the aim of developing a toolbox for the analysis of observational data. We consider three main context measures: attraction, compromise, and similarity. The proposed methodology hinges on *ex ante* calculation of context features for every alternative in multioption, multiattribute choice sets. This approach minimizes the computational complications of estimating the resulting choice model. The proposed approach is applied to air traveler choice data using an extensive observational dataset. This yields the first examination of all three context effects simultaneously in a large observational dataset. We discuss the consequences of product (re)design based on the results of the empirical exercise to showcase the potential use of the developed methodology in managerial practice.

1 | Introduction

The fact that behavioral biases exist in individual decision-making is well established (see Dowling et al. 2020 for a review of the evidence). One type of systematic departure from the classic utility maximization approach that seems particularly important is the effect of context (Trueblood et al. 2013; Köcher et al. 2019; Adler et al. 2024). The underlying theory posits that the context in which choices are made influences those decisions. Context effects appear to be persistent even under extensive deliberation in the choice process (Guo 2022; Kumar Padamwar et al. 2023). Choice context, in this literature, is the availability and the nature of choice alternatives (Tversky 1972; Huber et al. 1982; Simonson 1989).

Context effects have been systematically studied in psychology, marketing and information systems research (Kivetz et al. 2004; Rooderkerk et al. 2011; Frederick et al. 2014; Evangelidis et al. 2018; Mousavi et al. 2023). However, virtually all such studies have used simplistic experiments in controlled, compact settings. Namely, where decision makers are presented with limited options[1] and (very) few attributes across which these options differ. By contrast, most actual choices take place in much messier field conditions. This is especially true today, when much of our search and shopping activity has moved online. The proliferation of search engines allows each option to easily be compared with many alternatives across any number of distinct characteristics. Despite this, we know very little about the prevalence of context effects in these environments. This is problematic because context sensitivity of consumers has significant consequences for firms' optimal competition strategies (Apffelstaedt and Mechtenberg 2021).

There are two distinct streams of literature with a spirit similar to that of the current paper. On one hand, psychology researchers have been trying to extend original three-option, two-attribute studies to multioption, multiattribute settings. Scholars contributing to this stream have developed an array of theoretical models of cognition that allow for context effects (Roe et al. 2001; Bhatia 2013; Molter et al. 2022). These models have been empirically examined to varied extents, but always using laboratory experiments Trueblood (2022). Applying such multioption, multiattribute models to data usually requires slightly altering a choice menu (typically by adding an option or changing a characteristic of one of the options) and observing choices under both conditions. This is clearly not possible with observational data. Following tests in the laboratory, there is a belief that "the empirical principles are potentially applicable to the more complex real-life choices as well" (Roe et al. 2001). These laboratory evaluations have resulted in mixed conclusions regarding context effects. While some studies have documented the presence of certain context effects, others have found null results or even reverse effects (i.e., when a statistically significant effect is observed in the direction opposite of the one hypothesized) for every major context effect in the literature (Noguchi and Stewart 2014; Cataldo and Cohen 2019; Spektor et al. 2022). The overall conclusion is that presence of a given context effect, or of its reverse, is very sensitive to the choice environment Trueblood (2022).

On the other hand, there have been recent attempts in computer science to incorporate context effects in discrete choice models applied to observational data (Chen and Joachims 2016; Pfannschmidt et al. 2022; Bower and Balzano 2020). However, the aim in these cases is increased prediction accuracy (Tomlinson and Benson 2021). As a result, incorporation of context effects takes the form of generalizing choice models to allow for a departure from strict rationality assumptions.[2] These proposed generalizations of estimated functional forms usually do not distinguish different types of context effects. Additionally, they often run into computational difficulties, i.e., the estimation process is NP-hard (Maragheh et al. 2020).

The primary goal of the present study is to introduce a methodology to examine context measures in the field without explicit experimental interventions. We propose measures of context in multioption and multiattribute settings with the distinct goal of applying them to observational data. By doing so, we answer the recent call for unifying together fragmented streams of research on context effects contained within disciplinary boundaries (Evangelidis et al. 2024). Following Rooderkerk et al. (2011), we consider three context effects: attraction, compromise, and similarity. Attraction effect refers to the increase in attractiveness of a set of options as a result of adding an alternative to the choice set, compromise effect refers to the inclination of consumers to prefer options that represent a compromise across extreme sets of alternatives, and similarity effect refers to the drop in choice likelihood of an alternative once another similar alternative has been added to the choice set. Each of these effects occur in contexts that require a specific approach for making measurements applicable to observational data. Each context measure that is specific to a given alternative in a given menu is calculated prior to choice estimation. This avoids acute computational problems. Following the measurement of context, introducing measures in a discrete choice model enables the identification and estimation of context effects using observational data.

After describing the methodology, we present an application to a specific case from the field. We use an extensive observational dataset on airfare choice. In this setting, we document the attraction and similarity effects influencing choices in air travel booking data. We also detect a reverse compromise effect that seems to indicate that air travelers consistently prefer extreme alternatives (e.g., the cheapest or shortest flight) to alternatives that constitute a compromise among extreme options. In the concluding part of the paper, we discuss managerial implications for product redesign based on the results of the empirical exercise.

2 | Context Effects and Choice Modeling

The literature distinguishes three main context effects: attraction, compromise and similarity (Rooderkerk et al. 2011), the underpinnings of which are based on the perceptual framing of the choice problem. The attraction effect refers to a situation where the presence of an alternative A in a menu increases the attractiveness of a similar but strictly better alternative B (Huber et al. 1982). Individual's perceptions are manipulated using the alteration of the choice set such that we either increase the range of the dimension in which the target option is inferior to alternatives or increase the frequency of items along the dimension in which the target is superior to the alternatives (or both). The increase in the range by adding an alternative A is expected to decrease the importance of the difference between the target B and its competitors. On the other hand, increasing the frequency of items is expected to increase the weight of this dimension in consumer's decision-making process. In the literature, option A is often referred to as a decoy, and the attraction effect is sometimes called the decoy effect or asymmetric dominance effect (Evangelidis et al. 2018; Mousavi et al. 2023). In this context, superiority of option B over option A requires option B to be better than option A in at least one attribute and not to be worse in any of the attributes. While the attraction effect has been documented in various experimental settings (Huber and Puto 1983; Lehmann and Pan 1994; Doyle et al. 1999; Marini et al. 2023), researchers have also questioned the practical significance of the effect, because it requires clear identification of dominance relationships, which is not always obvious in real-life choice contexts (Frederick et al. 2014; Lichters et al. 2015; Rafaï et al. 2022). Two recent papers have studied the presence of attraction effect in the field using observational data. Wu and Cosguner (2020) examine an established precious stone retailer offering a wide selection of diamonds. They demonstrate that upon the detection of decoy alternative in the menu, consumers are several times more likely to purchase the attractive stone. Fridman et al. (2024) examine the data from a marketplace for digital freelance services and document a strong attraction effect. They also identify multiple important moderators for the effect, thus shedding light on mechanisms through which the attraction effect manifests itself.

The compromise effect refers to a phenomenon where consumers prefer choosing options in the inner portions of the choice set rather than those at the extremes (Simonson 1989). In other words, the presence of extreme options in the choice set could increase the choice probability of non-extreme alternatives. The original theory behind the compromise effect is focused on the decision-maker's uncertainty about the relative importance of features, which induces aversion to extreme options (Simonson 1989). Later research has shown that the compromise effect could also be derived from a model where consumers preference orderings are described by prospect theory (Kahneman and Tversky 1979) type value functions (value-shift models) (Bodner and Prelec 1994),[3] emergent-value models that are based on processing configural information about available options (Wedell and Pettibone 1996), as well as sequential search models with attribute-specific uncertainty (He 2024). The empirical literature has extensively demonstrated the presence of the compromise effect in laboratory settings (Dhar et al. 2000; Drolet 2002; Chernev 2004; Mao 2016; Guo 2022). Laboratory studies have also enabled the comparative study of multiple possible mechanisms driving the compromise effect Pettibone and Wedell (2000), and the study of the boundary conditions to the occurrence of the effect Evangelidis et al. (2018). Pinger et al. (2016) documented the compromise effect in the field using data from a restaurant over multiple menu iterations. Despite multiple factors that could potentially weaken the effect, the authors find a compromise effect in a meal choice setting.

The similarity effect refers to a circumstance where the likelihood of choosing the focal alternative decreases with the increasing number of items in the choice set that are similar to that focal alternative (Tversky 1972). The theory behind this effect posits that consumers search by iteratively eliminating options as they sequentially consider various factors and is related to the creation of a consideration set (Moe 2006); that is, a smaller set of choices that the consumer focuses on for further examination. Under this elimination by aspect process, the consumer gradually decreases the size of the consideration set while increasing its homogeneity. Such a process treats the choice as identifying and discarding worse alternatives rather than identifying the best ones. This process increases the probability that at each stage of elimination, similar alternatives receive similar treatment (by being either eliminated or selected into a smaller consideration set), this is the similarity effect. Like attraction and compromise effects, the existence of the similarity effect has been demonstrated by many experimental studies (see Wollschlaeger and Diederich 2020 for a review).

Over the years, multiple empirical models have been developed to model choice context. Empirical approaches usually model the choice context either in the structural part of utility function or in the error covariance part of the estimation process (Kamakura and Srivastava 1984; Dotson et al. 2018). Some of these models have the capacity to account for multiple effects at the same time (Tversky and Simonson 1993; Orhun 2009). These models extend a classical random utility model (McFadden 2001) in multiple directions using discrete choice modeling (Ben-Akiva and Lerman 1985). However, Rooderkerk et al. (2011) present a unifying model considering all three context effects. Instead of using advanced statistical techniques to remedy violations of the utility maximization assumptions associated with the existence of context effects (Luce 1959), their approach hinges on additive specification and *ex ante* calculation of three context measures for each item in the menu. The authors assume that choice estimator is additive in three context effects (along with a generic preference-driven part) and develop a methodology for quantifying three measures for each alternative prior to calculating the estimator. This is a particularly flexible approach that also ensures that researchers do not run into computational difficulties (such as NP-hard calculations). We follow this approach and formulate the utility that a consumer c attaches to an option i under a given menu m as being additive in two parts:

$$U^m_{c,i} = u_{c,i} + v^m_{c,i} + \epsilon^m_{c_i} \tag{1}$$

The first summand in this equation ($u_{c,i}$) denotes an inherent utility that the consumer c can derive from option i. This part depends only on the tastes of consumer c regarding the characteristics of option i. It is independent of other options contained in the menu. The second summand ($v^m_{c,i}$), denotes the context-dependent utility. The last summand is the random error term. We additionally assume that the context-dependent part of the utility can be represented as a linear combination of the three context measures,

$$v^m_{c,i} = a_1 Attraction^m_i + a_2 Compromise^m_i + a_3 Similarity^m_i \tag{2}$$

Thus, based on the utility formulation above, the three context measures (which are option- and menu-specific) need to be computed *ex ante*. Measures developed by Rooderkerk et al. (2011) are tailored to experimental data with a small number of alternatives in the choice set and a small number of attributes characterizing alternatives. This significantly limits the application of the unifying model of context effects. In the next section, we present a generalization of three context measures to a multioption, multiattribute environment that will further allow for the application of the unifying model to larger-scale observational data.

3 | Generalizing Context Measures

Generalizing context measures across many alternatives and attributes presents several challenges. The fact that the theoretical underpinnings of the three effects are diverse does not simplify the task. In the following subsections, we discuss specificities involved in the generalization of each measure. First, however, we concentrate on common challenges.

Conceptualizations of contextual effects hinge on the choice frequency comparisons between two alternatives. For example, in the case of the attraction effect, if adding a third alternative to a two-item menu induces some consumers to switch their choice from incumbent option A to incumbent option B, we could conclude that attraction effect is present. This is suitable for experimental setups where the researcher has complete control over the menus and can observe choices in both cases (with both a two-item menu and after a third item has been added). However, given that our aim is to generalize context measures to a wider range of situations and most importantly to observational data, it is necessary to take a more fine-grained view and quantify the context in which each of the alternatives is embedded. That would allow us to study the effect of the context on choice probabilities through inference across (very) different choice sets. Such an approach would be general enough to consider not only the addition of a new alternative to the menu but also any alteration of attributes for any of the items in the menu. For example, increasing the price of an alternative could decrease the probability of it being chosen. This would have a direct effect on the choice probabilities of other alternatives. However, the same price increase could also change the choice context and have additional effects on the choice probabilities of (at least some) alternatives.

Rooderkerk et al. (2011) take this approach in the simple case of two-attribute products. In the case of the attraction effect, the idea is to quantify how much attraction "power" a given menu provides to a given alternative. If option A dominates option B (i.e., is superior along at least some attributes and is not inferior in any of the attributes), while option B is not dominated by option C, the attraction power of A compared to C could be measured by the extent to which option A is better than option B. The more pronounced the dominance, the more pronounced the attraction effect. However, once we leave a neat context of three-item menus, option A could dominate not simply one but multiple alternatives at the same time. Consider the situation depicted in Figure 1. This menu has eight alternatives $\{A, B, C, D, E, F, G, H\}$, each of which is characterized with two attributes (V_1 and V_2, with a higher value of each attribute being more desirable for consumers). In this example option F dominates two alternatives (B and C). To straight-forwardly extend the approach by Rooderkerk et al. (2011) and calculate the attraction power of alternative F, we could find the center (halfway distance in two dimensions) between the two dominated options and then measure the distance from that point to F. Such a measure would capture the difference between two choice sets $\{A, B, C, D, E, F, G, H\}$ and $\{A, B, C', D, E, F, G, H\}$. In the latter case, F's attraction power is lower as option C' is closer to F than option C. However, such a measure would not accurately capture the difference between scenarios $\{A, B, C, D, E, F, G, H\}$ and $\{A, B, C'', D, E, F, G, H\}$. Under the latter case, F only dominates one alternative (B), so the setting changes qualitatively. While such qualitative differences are avoided in experimental settings by design, they are pervasive in observational data.

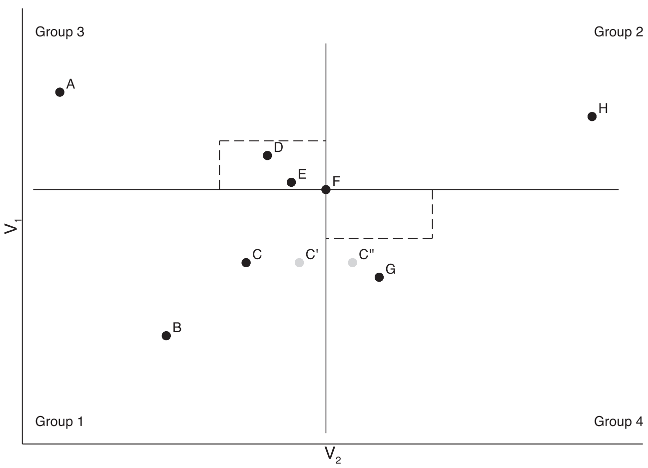

FIGURE 1 | Visualisation of the three context effects.

While we acknowledge that the move from C to C' does change the choice context, we argue that the context change is more pronounced in case of the move from C to C'' (or from C' to C'' for that matter). Even though the ideal measure would combine the features of the number of dominated alternatives and some measure of average distance between the focal alternative and the group of dominated options, in this paper, we take the approach of concentrating on the former because it is likely to have a more pronounced impact.[4] As a result, our approach will capture the context change between $\{A, B, C, D, E, F, G, H\}$ and $\{A, B, C'', D, E, F, G, H\}$ or $\{A, B, C', D, E, F, G, H\}$ and $\{A, B, C'', D, E, F, G, H\}$, but will evaluate no context difference between $\{A, B, C, D, E, F, G, H\}$ and $\{A, B, C', D, E, F, G, H\}$ (when focusing on option F). In what follows, the same approach is applied to similarity and compromise measures.

Once we move toward choices which have multiple attributes, it becomes apparent that there are two distinct types of choice characteristics that our measures could handle. One type of attribute constitutes product characteristics over which preferences are fairly similar for all customers. These attributes can easily be ordered from most to least preferred, using basic economic theory. The most obvious such characteristic is price. We can safely assume that every customer would prefer obtaining a given product for a lower price. We refer to such product characteristics as vertical attributes, which can usually be represented using numeric values. Previous work measuring context effects only considers vertical attributes. This is a requirement for defining preferential relationships that are necessary for identifying context effects. We adopt the same approach and consider only vertical attributes.

At the same time, menu items can also have another set of attributes for which there is no obvious, homogenous ordering. Consider color. We have no theoretical ground to assume that all consumers would prefer a car that is blue over one that is green, with all other attributes held constant. We call these horizontal attributes. Potential heterogeneity across decision makers in ordering in multiple categories in such attributes makes the inclusion of such features in the calculation of context effects impossible. In experimental settings, these attributes are often constant across treatments to prevent any confounding effects. However, in the field this usually cannot be done. Therefore, the study of context effects with observational data requires controlling for them statistically.

3.1 | Attraction

Previous studies of the attraction effect concentrate on carefully designed small choice sets in experimental settings (Huber et al. 1982; Huber and Puto 1983). In such settings, an alternative is added to the choice set in a position such that it is unequivocally inferior to (only) one of two items already present in the menu. Note, again, that identification of inferiority requires the attribute under consideration to be vertical. This manipulation introduces an asymmetry between the two incumbent alternatives—one alternative now dominates the decoy, while the other does not. The attraction effect implies that such manipulation increases the attractiveness of the dominant incumbent option with respect to the other incumbent alternative.

A standard measure of the attraction effect considers a tradeoff between two (vertical) characteristics. Consider $i \in N$ vertical attributes—V_i—for a set of two options A and B. In two dimensions ($N = 2$), we start out with $V_1(A) > V_1(B)$ and $V_2(A) < V_2(B)$ and then introduce an alternative C such that $V_1(A) > V_1(C) > V_1(B)$ and $V_2(C) < V_2(A) < V_2(B)$. Under such circumstances, C is dominated by A but not by B. This introduces asymmetry to consumer considerations and increases the probability that the consumer will choose option A. Generalizing this concept to multiple (vertical) attributes is straightforward. For $N > 2$, we again start out with A being preferred over B in some ($j > 0$) dimensions, while B is preferred to A in some others ($k > 0$), such that $j + k \leq N$. Then, we need an alternative C that will be strictly worse than A in at least one dimension while not being better in any other dimension and being better than B in some dimensions while being worse in some others. As long as these two conditions are satisfied, the attraction effect suggests that presence of C will result in boosting the choice probability of A.

Generalizing this approach to multiple alternatives is somewhat more challenging because rather than one comparison (A vs. B in the case above) for a choice set with M alternatives, there are $M(M-1)/2$ potential comparisons to consider. Under real-life circumstances, it is easy to identify situations where more than one of $M(M-1)/2$ relationships carry the potential for attraction effect. In addition, for any given pair of choices, we could have multiple decoy options generating attraction effects. The final complication is that option A may have one set of decoy alternatives and option B another set of decoys. In these contexts, it is not clear which option the attraction effect favors.

We propose that the difference in the power of attraction between two alternatives generated by the menu should be proportional to the difference (i.e., asymmetry) in the number of alternatives they each dominate. Interestingly, this could be achieved by calculating the number of dominated alternatives by each member of the menu and estimating how this quantity influences choice likelihood. Consider the different positions that option C can take with respect to options A and B. If option C is neither superior (dominant) nor inferior (dominated) to any of the options (A and B), or if it dominates both of the focal options, then it cannot generate an attraction effect for either A or B. If option C is dominated by both of the options in a focal pair, it generates an attraction effect for both of them. However, the attraction power it yields to each of the options is the same; thus, it does not act as a discriminant across the two options in question (A and B). In all of these cases, option C's location contributes similarly to the choice probability of both options (A and B).

Finally, if option C is dominated by only one of the two focal alternatives, it generates a discriminatory attraction effect favoring option A and increasing its probability of being chosen. As a result, the number of options that the current alternative dominates in a menu (appropriately normalized by the menu size for a comparison across different choice settings) measures the (relative) extent of the attraction generated by the menu. For example, contrast the probability of choosing option F vs. G in Figure 1 across two sets of menus $\{A, B, C, D, E, F, G, H\}$ and $\{A, B, C'', D, E, F, G, H\}$. The choice probability is higher in the former situation (where F dominates two alternatives while G dominates one) than the latter case (where F and G each dominate one alternative). Even though in these cases both alternatives do have some attraction relative attraction of option F compared to option G is stronger in the former scenario than in the latter. Therefore, we measure attraction favoring the focal option F as

$$Attraction(F) = \frac{1}{M} O(Dominated) \tag{3}$$

where $O(Dominated)$ measures the number of alternatives in the menu that the focal option (F) dominates, and M is the menu size. Normalizing the measure by the menu size makes it comparable across menus of different sizes.

It may appear that simply counting the number of options dominated by an alternative is not a proper measure of the power of the attraction, since it does not explicitly consider the requirement of asymmetric dominance. However, note that the difference in the number of dominated options between two alternatives equals the difference in the number of decoys targeting those two alternatives, since options that are dominated by neither alternative are not counted, and options dominated by both alternatives are counted for both. Therefore, the parameter a_1 in Equation (2) can be interpreted as the utility premium of having one more decoy over the competing alternative (multiplied by the menu size). We expect that the higher the $Attraction(F)$ measure in favor of the focal option (F), the higher the choice probability of F, *ceteris paribus*.[5]

3.2 | Compromise

The compromise effect is traditionally understood and operationalized in a three-option, two-attribute setting (Simonson 1989; Dhar et al. 2000). These two attributes need to be vertical so that we can define universal preference relationships. Let us consider the similar starting situation of options A and B as in the previous sub-section: $V_1(A) > V_1(B)$ and $V_2(A) < V_2(B)$. The addition of option C to this menu such that $V_1(C) > V_1(A) > V_1(B)$ and $V_2(C) < V_2(A) < V_2(B)$ makes option A a compromise between two extreme options (B and C). The compromise effect maintains that such an alteration of the menu would disproportionately benefit alternative A compared to alternative B.

To formulate the general measure of the compromise, let us first consider the case of multiple options (M) in two dimensions (attributes, $N = 2$). We visualize the compromise calculation over multiple options in Figure 1 with $M = 8$ case; we consider menu $\{A, B, C, D, E, F, G, H\}$. To quantify the extent of the compromise that focal option (F) carries in this menu, we propose to split all other $M - 1$ alternatives into four groups. Let group 1 contain all alternatives for which $V_1(i \in Group_1) \leq V_1(F)$ and $V_2(i \in Group_1) \leq V_2(F)$. These are the alternatives dominated by the focal option. In the menu in Figure 1, this set contains options B and C. Let group 2 contain all alternatives for which $V_1(i \in Group_2) \geq V_1(F)$ and $V_2(i \in Group_2) \geq V_2(F)$. All these options dominate the focal option. This set contains option H in Figure 1. Our focal option cannot constitute a compromise between any pair of alternatives that is included in any of these first two groups of alternatives. Next, let group 3 contain all alternatives for which $V_1(i \in Group_3) > V_1(F)$ and $V_2(i \in Group_3) < V_2(F)$ and group 4 contain all alternatives for which $V_1(i \in Group_4) < V_1(F)$ and $V_2(i \in Group_4) > V_2(F)$. As Figure 1 shows, group 3 contains options A, D and E, while group 4 contains option G. Our focal alternative can be viewed as a compromise between groups 3 and 4. As the quantification of the extent of such a compromise, we define

$$Comp(F) = \frac{min\left(O(Group_3); O(Group_4)\right)}{max\left(O(Group_3); O(Group_4)\right)}\left(O(Group_3) + O(Group_4)\right) \tag{4}$$

where $O(Group_j)$ measures a number of alternatives in group j . The first multiplier (the ratio) in the measure quantifies the asymmetry across the sizes (in terms of number of alternatives) of the two groups, while the second multiplier (the sum) quantifies the combined size of the two groups across which the focal option is a compromise. For option F in Figure 1, this value is $Comp(F) = \frac{1}{3} \times 4 = 1.33$. If either of the two concerned groups is empty, the value is zero, corresponding to the fact that the focal alternative is at the extreme edge of one of the dimensions and therefore is not a compromise. As a result, our measure will be strictly zero for options A, B, G, and H. On the other hand, the better the balance between the size of the two groups, the more valuable the compromise that alternative F provides. So, the same measure for option E in Figure 1 is four. Alternative E also corresponds the compromise between four alternatives (like option F), but the comparison groups are better (and in this case, perfectly) balanced. Note that the measure is also increasing in the number of total options in two comparison groups and that the same measure for option C is two, even though (similar to option E) it exhibits prefect balance across the sizes of two comparative groups. This reflects the fact that option E is a compromise between larger sets of extreme alternatives.[6]

Extending the compromise measure to multiple dimensions is somewhat more challenging because increasing number of vertical dimensions increases the number of different ways a given option can be a compromise. The $N = 2$ case has one pair of groups to compare. In the case of $N = 3$, however, a focal alternative can be a compromise between multiple pairs of option groups. For example, option F can be a compromise between two groups Z and Y such that all options in group Z are superior to option F in dimensions 1 and 2 but inferior in dimension 3, while options in group Y are inferior to option F in dimensions 1 and 2 but superior in dimension 3. Permutation calculus guarantees that there are three such potential comparisons. However, this is not all. Option F can also be a compromise between two groups X and W such that all options in group X are superior to option F in dimension 1 but inferior in dimension 2, while options in group W are inferior to option F in dimension 1 but superior in dimension 2, as long as dimension 3 is constant across all options in groups X and W, as well as F. Permutation calculus guarantees an additional three such comparisons.

As a result, moving from two to three dimensions increases the number of potential comparisons for calculating compromise value of an alternative from one to six. Appendix A derives the number of comparison alternatives necessary to cover all potential ways a focal alternative can be a compromise as a function of the number of dimensions. However, as all the above-defined groups are mutually exclusive, generalization of the compromise measure in N dimensions would require summation of our comparison group-specific compromise measure over all comparison groups and consequent normalization with respect to menu size. Thus,

$$Compromise(F) = \frac{1}{M}\sum_j Comp(F)_j \tag{5}$$

where j runs over all possible comparison groups. Given our compromise measure, we expect that the higher its value, the greater the choice probability of the focal alternative.

3.3 | Similarity

Operationalizing the similarity measure across three options and two vertical dimensions is straightforward (see Rooderkerk et al. 2011). Increasing the menu size beyond three options introduces an important challenge of defining options that are similar to the focal alternative and those that are not similar to it.

The sufficient condition for quantifying the similarity effect requires detecting the number of other alternatives that are similar to the focal option. Similarity between two alternatives can be defined as the circumstance when values of all vertical attributes are sufficiently close across two options. This necessitates the definition of a cutoff point that would qualify two options as sufficiently close. There could be many approaches to this, but we argue that appropriate definition should be both menu- and attribute-specific. We propose to set a cut-off point as a certain percentage (r) of the range (R) within a dimension in a given menu. Consider the situation of menu $\{A, B, C, D, E, F, G, H\}$ in Figure 1, under the condition that the range displayed range on the two axes is the same. In this example, the range of dimension V_1 (R_1) is smaller than the range of dimension V_2 (R_2). Therefore, it would be natural to expect that two options equidistant from each other in these two dimensions could be considered similar in V_2 but not in V_1. Once we set the cut-off value of r, which is constant across dimensions because R_n is read from the menu itself for every dimension, we can create a perimeter around a given option that would capture all options similar to the focal option. Note, however, that options from groups 1 and 2 cannot be included in the count of similar options because they have dominant/dominated relations with the focal option (Bergner et al. 2019). This leaves us with boundaries visualized by dashed rectangles around option F in Figure 1. Every option (in this case options D and E) within this area qualifies as an option similar to F. After counting all options similar to F, we can calculate

$$Similarity(F) = \frac{1}{M}O(Similar_F) \tag{6}$$

where $O(Similar_F)$ counts the number of options for which $|V_n^F - V_n^i| \leq rR_n,\ \forall n$. We use the value of $r = 0.2$ in the visualization in Figure 1.[7] Given Equation (6), we expect that the higher the value of the similarity measure, the lower the purchase probability of the focal option, *ceteris paribus*.

It is worth mentioning that by modeling context effects we do not discount the importance of more rational reasons as drivers of consumer choices. It is absolutely essential that estimation strategy takes those reasons into account. Modeling those rational reasons would require estimating effects of all vertical variables, at the very least. It is, however, important to highlight that all our context measures do also depend on the same vertical variables (albeit used only jointly). Therefore, our approach to model context effects rehashes the same information that is necessary to model rational choice reasons. Our approach to taking into account the context pays closer attention to how vertical variables jointly contribute to consumer choice beyond individual vertical variables that account for rational reasons for choice.

4 | Empirical Application

In this section, we present an empirical application using the generalization of the three context measures developed in the previous section. This application concerns the analyses of a large-scale observational airfare choice dataset. This is a highly heterogeneous dataset containing choice setups that vary in terms of number of alternatives, as well as across origin-destination city pairs. Using the information about available menus, we calculate three generalized context measures for each option and then use those measures in a discrete choice analysis.

Our observational dataset involves the merger of two sources. The first dataset constitutes a list of all bookings made in Europe on intra-European routes between December 2013 and June 2014, extracted from the Marketing Information Data Tapes database. In addition to all the booking details (e.g., number of passengers, price), it contains the timestamp of booking and the identity of the booking office (all offline and online outlets have unique identifiers). The second source of data contains information on all air travel searches performed on one of the most comprehensive air travel booking services operated by Amadeus S.A.S. This dataset also contains trip specifics and the identifier of the office where the search was performed. Most importantly, the latter dataset contains information on all possible alternatives that could have been presented to the traveler at the time of the search, but does not contain information on which option, if any, the traveler chose. Matching these two datasets across office identifier, search/booking time, trip origin and destination, trip dates and number of passengers results in a merged dataset that allows us to identify chosen itineraries within the option menus delivered during the search.[8] An important limitation of the data is that we have no way of ensuring that the consumer has actually seen an exhaustive list of alternatives available at the time of booking. We do know, however, all the options a customer might have seen. Even though this is a drawback for a researcher, this is a standard experience for the practitioner (e.g., a manager or a recommender system designer). Practitioners designing recommender systems need to create algorithms based on the set of existing alternatives without much visibility on the subset of options a particular user will be interested in or will eventually see.

The matched dataset (previously used by Mottini and Acuna-Agost 2017 and Mirzayev et al. 2021) consists of about 13,000 choice sessions with around one million choice alternatives in total. Every alternative is a round-trip flight and has a number of attributes, including ticket price, date and times of all inbound and outbound flights, number of flights in the itinerary, number of airlines, number of days before travel and a few other less important attributes.

Menus with only one available alternative do not allow the consumer to make a choice; and menus with two alternatives do not allow for the calculation of context measures. As a result, menus of those sizes are discarded. Data on choices contains at most 100 alternatives for each choice session, even if more choices potentially existed. As a result, our data are truncated from the right. This creates a large number of menus of exactly 100 alternatives, some of which may be incomplete. To deal with this issue, we simply confine our analysis to menus having between 3 and 99 alternatives. In the end, we are left with a dataset with 5784 choice sessions with 344,386 alternatives (with an average menu size of about 60).

Descriptive statistics of the main variables are given in the upper portion of Table 1. The first four variables on the list are the attributes designated as vertical in the choice process. For the purposes of this paper, we assume that consumers prefer lower values in each case, all consumers prefer lower prices, shorter trips, fewer layovers and not having to change airlines too frequently. The data also contains two sets of horizontal attributes,

TABLE 1 | Descriptive statistics of the European airfare booking dataset.

Variable	# of obs.	Mean	SD	Min.	Max.
Price (in EUR)	344,386	683.98	2272.36	59.55	86,997.00
Trip duration (in minutes)	344,386	9447.20	20,527.56	80.00	432,970.00
Number of flights	344,386	2.97	0.98	2.00	6.00
Number of airlines	344,386	1.27	0.46	1.00	5.00
Menu size	344,386	73.91	20.19	3.00	99.00
Attraction	344,386	0.25	0.24	0.00	0.99
Compromise	344,386	0.04	0.07	0.00	0.80
Similarity	344,386	0.04	0.05	0.00	0.73

the departure times and dates of outbound and inbound flights. We treat these attributes as horizontal as we have no clear way of defining consumer preferences about them. Even though horizontal attributes do not go into the quantification of the context, it is important to control for them at the stage of choice modeling. We also have three attributes that do not vary across alternatives within each menu. These are the number of days between when choice was made and the start of the trip, whether the trip was domestic or international, and whether it is intercontinental.[9] We also use these attributes as controls in our choice model to ensure that different choice instances are comparable to one another. Additionally, we have the menu size, which constitutes an important feature of the choice context.[10]

Before getting into the data analysis, to eliminate potential scale effects, we performed z-score normalization on vertical attributes following $Z = \frac{x-\overline{x}}{\sigma}$, where $\overline{x}$ is the mean and σ is the standard deviation of variable x.

4.1 | Measurement of Context

To measure the option-specific context, we follow the methodology outlined in the previous section. For attraction, we count the number of alternatives dominated by a given option within the menu. This is implemented across all four vertical attributes. The compromise is also measured across all four dimensions. This results in 25 pairs of comparison groups for each alternative (see Equation A1 in Appendix A). For similarity, we use a cutoff value of $r = 0.2$ (or 20%).[11]

The lower portion of Table 1 presents descriptive statistics of the three context measures in the dataset. To have a more fine-grained feel of the measure, Figure 2 plots each of the context measures along the different menu sizes in the dataset. Significant variation in menu sizes is a distinguishing feature of our analysis. Previous context effect study designs have held this feature constant or nearly constant. Having menus of different sizes allows us to draw conclusions about the dependence of context effects on menu size, which is an important characteristic in the field. It is worth noting that compromise and similarity measures seem to be increasing in menu size for smaller menus. This is not surprising, because smaller menus limit the number of opportunities a focal option can be similar to others. Smaller menus also constrain the focal option from being a compromise among larger groups of alternatives. At any rate, even though all three measures are normalized by menu size, we still see a dependence of the resulting context measures on menu size, which calls for careful modeling of this feature in the discrete choice analysis.

4.2 | Choice Modeling

To examine the context effects on choices in the airline booking data, we estimate probit models augmented by our context measures. We estimate a set of random effects probit models to estimate the choice likelihood of each alternative in a given menu. Random effects models provide the desired flexibility in estimation procedure (i.e., menu-specific intercepts and slopes) for the application at hand. Beyond flexibility, an important advantage of random effects models is that we can keep menu size as a meaningful modeling component. In an alternative fixed effects specification, the menu size would be eliminated because it does not vary across options within the same menu. We do, however, verify that the results from fixed effects estimation are consistent with the ones obtained by the random effects fitting.[12] An important advantage of probit model for studying context effects (over, for example, logit) is the feature that probit does not explicitly require the assumption of the independence from irrelevant alternatives (IIA). If our (saturated) model perfectly accounts for all context effects this would not be a concern. However, as we cannot guarantee that human choices are not affected by any other context features (that have not been accounted for by our model or have not yet been hypothesized), having no implicit assumption of IIA is an additional advantage. It is worth noting that the results are not sensitive to alternative model specifications (see the Appendix H (Table H1) for the results from logit estimation).

Before getting to the estimation, we need to transform our departure time variables in order to meaningfully include them in the estimation procedure. To make departure time information as tractable as possible, we generate a set of variables. First, we generate a day of the week variable for the outbound flight. Second, we generate a variable that measures the duration of the stay at the destination.[13] These two variables together describe inbound and outbound flight timing characteristics at the level of one day. However, consumer preferences might be defined on a smaller scale. Therefore, we also generate two variables that describe the exact time of the day of outbound and inbound flights. These variables, $t_{out}, t_{in} \in [0;1)$, are measured as a fraction of a day,

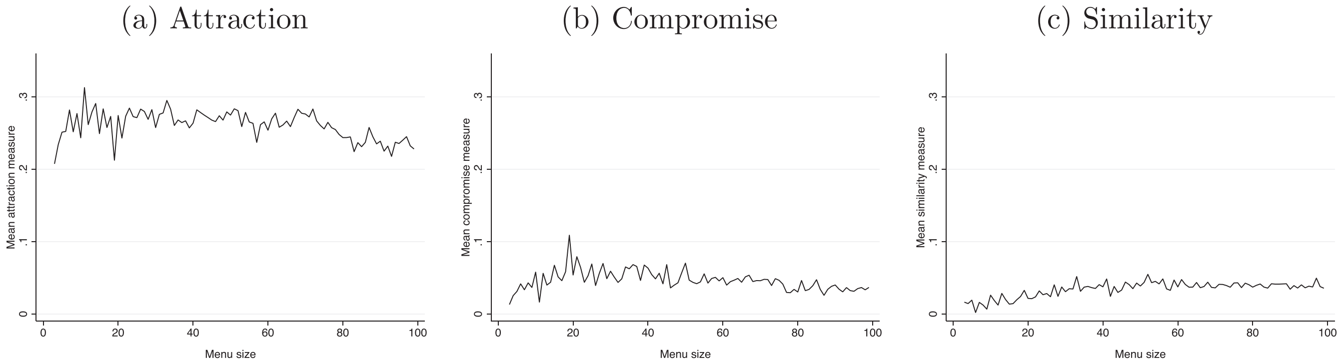

FIGURE 2 | Averages of three context measures across varying menu size.

TABLE 2 | Estimation results.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Price	−0.237***	−0.245***	−0.230***	−0.243***	−0.239***	−0.219***
	(0.007)	(0.007)	(0.009)	(0.007)	(0.008)	(0.009)
Trip duration	−0.092***	−0.111***	−0.089***	−0.111***	−0.113***	−0.084***
	(0.007)	(0.007)	(0.011)	(0.007)	(0.007)	(0.011)
Number of flights	−0.364***	−0.322***	−0.324***	−0.319***	−0.305***	−0.306***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)	(0.007)
Number of airlines	−0.209***	−0.210***	−0.203***	−0.209***	−0.205***	−0.195***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)
Menu size	−0.016***	−0.016***	−0.016***	−0.016***	−0.015***	−0.015***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Attraction			0.111**			0.147***
			(0.042)			(0.042)
Compromise				−0.284**		−0.171
				(0.094)		(0.094)
Similarity					−1.931***	−1.940***
					(0.187)	(0.190)
Constant included	Yes	Yes	Yes	Yes	Yes	Yes
Horizontal variables as controls	No	Yes	Yes	Yes	Yes	Yes
Number of observations	344,386	344,386	344,386	344,386	344,386	344,386
Number of choices	5784	5784	5784	5784	5784	5784
Akaike information criterion	46,151	45,268	45,264	45,260	45,146	45,134
Log likelihood	−23,068	−22,618	−22,615	−22,613	−22,556	−22,548

Note: Results from random effects probit estimations. Complete estimation outputs can be found in the Appendix I (Table I1). Groups represent menus in the datasets. Robust (menu-level clustered) standard errors are in parentheses. Significance key: "***" indicates $p < 0.001$, "**" indicates $p < 0.01$, "*" indicates $p < 0.05$.

such that $t_i = 0$ corresponds to the midnight, while $t_i = 0.5$ corresponds to midday. We further apply a cosine transformation to these variables: $departure\_time(t_i) = \cos(2\pi t_i)$. This confines the departure time variable to the interval $[-1; 1]$ and ensures that times right around midnight are similar to each other in terms of this variable. These transformations result in a total of four variables describing departure timestamps for an outbound and inbound flight pair, which is a horizontal attribute of the alternative.

We estimate a sequence of 6 models and present results in Table 2.[14] We use random effects Probit regressions with robust (menu-level clustered) standard errors.[15] We start out by fitting two simple baseline models of consumer choice. Model 1 is the simplest estimation and includes only the four vertical attributes and menu size as independent variables. Model 2 extends this model by adding four horizontal (and menu level-invariant) attributes. In both cases, with or without horizontal attribute controls, all vertical variables generate meaningful results. Consumers clearly have preferences for shorter, cheaper flights with fewer layovers and airline changes. Menu size also generates negative and significant coefficients indicating the fact that the probability of an option to be chosen is decreasing with increasing menu size.

To further extend model 2, we estimate three models (3 through 5), each of which incorporates one of the context measures. Finally, we estimate a saturated model (model 6) that incorporates all three context measures at once. Table 2 indicates consistency between the coefficient estimates of model 6 and those from models 3 through 5. This set of models also allows us to evaluate the effect of the three context measures on consumer choice. In line with the theory, we see the presence of attraction and similarity effects. More precisely, the increase in the attraction measure for a given option is associated with the increased likelihood of that option being chosen, *ceteris paribus*. Conversely, the increase in the similarity measure is associated with the decrease in the likelihood of the option being chosen. Both of these effects are statistically significant and are in the hypothesized direction.

Table 2, however, also indicates the existence of a reverse compromise effect. The compromise effect posits that if an option represents a compromise between extreme alternatives, it will have a higher likelihood of being chosen. By contrast, our results indicate that increasing our compromise measure is associated with the decrease in the likelihood of an option being chosen. This effect is statistically highly significant in model 2 but loses statistical significance in the saturated model 6 ($p = 0.089$). We can thus conclude that in the context of airfare choice, consumers prefer extreme options to those that represent compromise. The presence of the reverse compromise effect has been previously reported by Cataldo and Cohen (2019). The authors manipulated the appearance of the menu on computer screens in front of subjects in order to force them to make either by-attribute or by-alternative comparisons in the choice process. They reported that while by-dimension comparison elicited the compromise effect, the by-alternative comparison resulted in the reverse compromise effect. Unlike Cataldo and Cohen (2019), we detect a reverse compromise effect in the field setting in the context of air travel. We identify two potential explanations for this finding. One is related to the search process by the consumers. The other to consumer preferences in the context of air travel.

Consider a search process by the consumer (following exclusionary search using either by-alternative or by-attribute comparisons Heller et al. 2002) in a setup where we have many more options in the menu than there are attributes characterizing each alternative. Under such circumstances, it is unlikely that by-dimension or by-attribute search would result in a unique option to purchase. At some point, the consumer would need to perform some by-alternative comparisons to be able to identify the preferred alternative. As we increase the number of relevant dimensions, by-dimension comparison might make by-alternative comparison redundant because it becomes increasingly likely that the process converges to a unique alternative for choice. We argue that in a setup with a large number of options, if the number of relevant dimensions of the product is high (as in case of complex products), it is more likely that consumers complete their selection using by-dimension search. In such a situation, using Cataldo and Cohen (2019) findings, we are likely to observe the anticipated compromise effect. On the other hand, as we move toward more standardized products by reducing the number of relevant dimensions, the need for by-alternative comparisons increases. In this case it is more likely to observe the reversed compromise effect. Travel is a relatively standardized product with only a handful of relevant dimensions, which alignes our findings with those of Cataldo and Cohen (2019).

An alternative explanation of reverse compromise effect could be linked to consumer preferences in the specific context of travel, in which consumer preferences could be strongly anchored to one of the four vertical attributes. For example, if a traveler attaches a particular importance to price, she will be reluctant to choose an option that increases another (vertical) dimension over a cheaper option. This is understandable given the context of our empirical exercise; the two largest groups of air travelers are holidaymakers, who are price-sensitive and do not readily trade away cheaper prices for shorter travel time, and business travelers, who are time-sensitive and do not trade away flight duration for a decrease in price.

Table 2 presents the statistical significance of each effect. However, in order to gain an accurate feeling of how each of the context measures contributes to choice probability, it is necessary to look at marginal effects. Figure 3 depicts the marginal effects for all three context measures across different menu sizes. These are marginal probabilities of a positive outcome for a corresponding context measure. They are obtained from the saturated model (model 6) and are evaluated at mean values of all attributes except menu size, which is variable in the graph. Overall, we see that there is a negative association between context effects and the menu size, which is consistent with previous findings with smaller menus (Stanley and Wedell 2024). Panel (B) of Figure 3 shows that the compromise effect is not statistically significant at any menu size (with 95% confidence). As far as economic size of the effects is concerned, Figure 3 indicates that while the sizes of the attraction and compromise effects are comparable, margins for similarity are higher by an order of magnitude. This means that change in similarity measure is going to have a much larger effect on choice likelihood than the same-sized change in any of the other measures. It is also important to note that despite the fact that marginal effects of all measures decline with menu size, effects for attraction and similarity remain statistically significant even for large menus. This indicates that gaging context effects (especially attraction and similarity) remains a potent lever for influencing consumer choices through menu design for menus of any size.

4.3 | From the Menu to a Likely Consideration Set

An important drawback of our dataset is the feature that we have no way of knowing which options available on the market

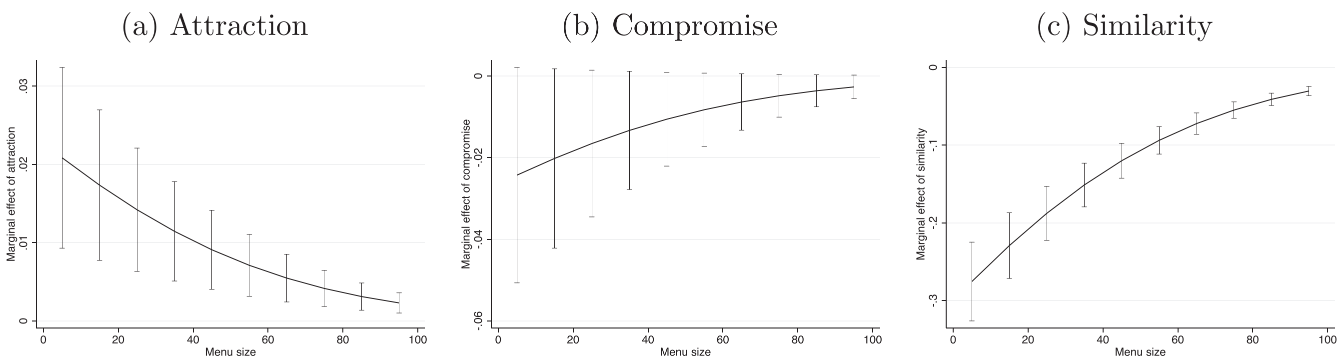

FIGURE 3 | Margins for the three context effects. Estimates derived from model 6 while holding all variables at sample means. Confidence intervals correspond to a 5% error rate.

actually reached a given consumer before the choice was made. This drawback, however, is not unique to our work. A previous study of contextual choices with observational data by Otto et al. (2022) constituted the choice menu from all available similar restaurants in a given location. One way of thinking about overcoming this problem is to consider a subset of the menu that constituted a consideration set of the consumer. Hard data on this point is impossible to obtain in observational settings. Each menu, especially one that contains many options characterized by many attributes, is likely to contain options that are not at all appealing to a given consumer and are therefore never taken into consideration. The task here is to identify and discard those options from the menu in order to refine the consideration set. The theoretical underpinnings of the consideration set formation are varied, and the application of different screening rules usually results in varied outcomes (Gilbride and Allenby 2004). Such screening rules are commonly based on certain threshold values for one or several characteristics of the product. Our approach is to use menu characteristics and the information about the eventual choice that the consumer made to eliminate some options from a potential consideration set. Given that we are not considering this exercise as a substitute to the approach presented in the previous section but rather a robustness check remedying a specific data problem, our approach does not need to be driven by theoretical considerations and frees us from making a choice among competing screening rules. We simply deduce which options are most likely to have been presented to the consumer given what we know about their preferences. In spirit, this is similar or perhaps complementary to the calculation of the similarity measure. However, it is necessary to include all available information about all options in the menu in such a procedure. This concerns both vertical and horizontal variables.

Clustering using machine learning gives us the opportunity to operationalize the reduction of a menu to a potential consideration set. Clustering algorithms are unsupervised machine learning techniques that require no explicit definition of similarity and can be applied to data containing both vertical and horizontal attributes. These algorithms use internally consistent evaluation criteria to partition the input group of objects into multiple sub-groups. By definition, items belonging to the same group are similar to each other, while items belonging to two different groups are dissimilar. Therefore, we argue that all options deemed similar to the chosen option by the machine learning algorithm would constitute a likely consideration set of the consumer.

Some popular clustering algorithms, like K-means clustering (Lloyd 1982), require pre-specification of the number of subgroups (clusters) into which a researcher would like to split the collection of objects. The use of such algorithms would be problematic in our situation as we have no *a priori* idea about the suitable number of clusters in each menu. Another class of algorithms automatically calculate the optimal number of detected clusters, given the specification of the similarity measure. A prominent representative of this class is the affinity propagation (Frey and Dueck 2007) procedure. Appendix D provides a short sketch of the essence of the affinity propagation algorithm. Being able to automatically detect a suitable number of clusters is necessary for our application. A consequence of this feature is that the size of each cluster—and most importantly of the cluster to which a chosen option belongs, a likely consideration set—need not be specified in advance and will be decided by a machine learning algorithm based on the menu characteristics. As a result, we employ affinity propagation for detecting a likely consideration set within each menu.[16]

In what follows, we conduct a choice modeling exercise from Section 4.2 on the reduced dataset, which only contains options belonging to the same cluster as the chosen alternative in every menu. Additionally, we eliminate menus where the cluster containing the chosen alternative is not larger than two options (there are 767 such menus). Reducing the dataset eliminates nearly 88% of observations that are deemed to be out of likely consideration sets of the consumers. This implies a reduction of the average menu size from about 60 to about 8. The exercise requires re-normalization of our vertical variables, re-calculation of three context measures and re-fitting of statistical models. Table 3presents the results from random effects probit estimations. The results are consistent with the original estimates from Table 2, with the compromise measure achieving statistical significance even in the saturated model 6. Figure 4 shows much more consequential marginal effects for attraction and compromise measures compared to those from Figure 3, indicating that context effects are much more pronounced once we discount for the noise coming from irrelevant alternatives.[17]

An alternative approach to using information from clustering algorithm is to account for the difference in the relationship between two options belonging to the same cluster and those belonging to different ones in the process of calculating our context variables. In this complementary exercise, instead of discarding all options that do not belong to the cluster of the chosen alternative, we create two sets of context measures. One is calculated based only on the options that belong to the same cluster to which the focal alternative belongs (i.e., inside the cluster measures). The other is calculated based only on the part of the menu that does not belong to the same cluster as the focal alternative (i.e., outside the cluster measures). Given these two sets of context measures, we estimate the extended model 6 with random effects probit procedure. Figure 5 plots estimated coefficients across two sets of measures. As one would expect, we see measures inside the cluster having much larger magnitudes than those outside the cluster. Outside the cluster measures for attraction and compromise are statistically insignificant. This approach also allows us to statistically test the difference between the coefficients inside and outside the cluster for each pair of measures. Figure 5 also reports *p* values of the three statistical tests. While the differences between the coefficients inside and outside the cluster are statistically significant for attraction and compromise measures, the statistical test for the pair of similarity measures fails to achieve statistical significance. This, however, could be explained by the fact that clustering coefficient and similarity measure have a significant overlap.

As already noted, our data have two disadvantages. First, we do not see exactly which options were examined by the consumer. Second, we cannot identify multiple choices made by the same consumer. As a result, despite estimating random effects models, our analyses might be polluted by variation in consumer-level characteristics. These features are not

TABLE 3 | Estimation results using the reduced dataset.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Price	−0.123***	−0.129***	−0.089***	−0.128***	−0.131***	−0.091***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Trip duration	−0.032***	−0.053***	−0.046***	−0.049***	−0.054***	−0.044***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Number of flights	−0.002	−0.024***	−0.034***	−0.021***	−0.024***	−0.035***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Number of airlines	−0.016*	−0.018*	−0.022**	−0.017*	−0.023**	−0.025**
	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Consideration set size	−0.061***	−0.060***	−0.064***	−0.059***	−0.059***	−0.063***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Attraction			0.761***			0.732***
			(0.033)			(0.034)
Compromise				−0.695***		−0.457***
				(0.092)		(0.093)
Similarity					−1.080***	−0.985***
					(0.141)	(0.140)
Constant included	Yes	Yes	Yes	Yes	Yes	Yes
Horizontal variables as controls	No	Yes	Yes	Yes	Yes	Yes
Number of observations	40,916	40,916	40,916	40,916	40,916	40,916
Number of choices	5017	5017	5017	5017	5017	5017
Akaike information criterion	28,771	28,674	28,176	28,611	28,637	28,122
Log likelihood	−14,379	−14,321	−14,071	−14,289	−14,301	−14,042

Note: Results from random effects probit estimations. Complete estimation output can be found in the Appendix I (Table I2). The reduced dataset is generated after discarding all alternatives not belonging to the same cluster as the chosen alternative in each menu, as well as those menus where potential consideration sets do not contain at least three alternatives. For comparison, Appendix I (Table I3) refits the regressions from Table 2 to the dataset containing only menus included in the estimation of results in this table. Clustering performed using affinity propagation algorithm. Groups represent menus/clusters in the datasets. Robust (menu-level clustered) standard errors are in parentheses. Significance key: "***" indicates $p < 0.001$, "**" indicates $p < 0.01$, "*" indicates $p < 0.05$.

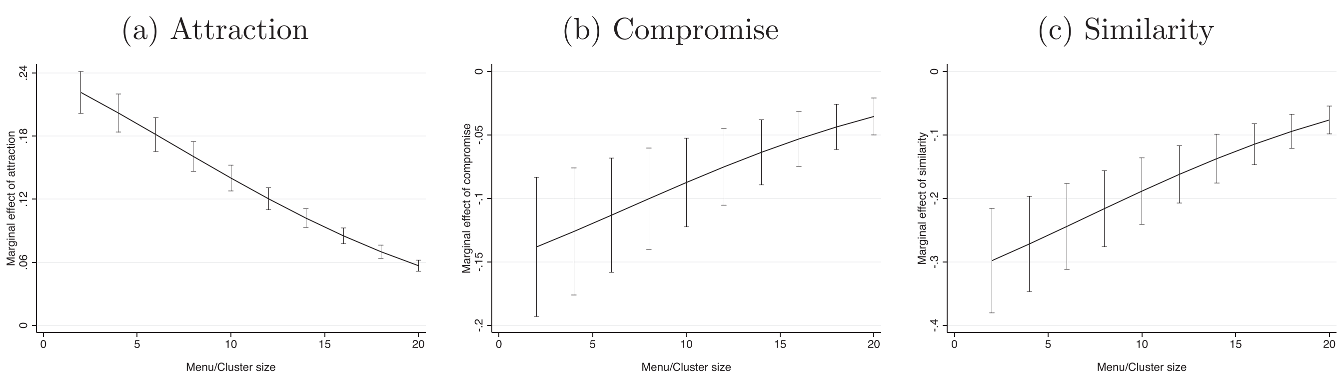

FIGURE 4 | Margins for the three context effects for a smaller consideration set. Estimates derived from model 6 estimated using data containing only options that belong to the same cluster as the chosen alternative in each menu. Confidence intervals correspond to a 5% error rate.

uncommon in observational data, and it is important that the proposed methodology be nimble enough to deal with such shortcomings. Such issues, however, do not exist in the experimental data. Naturally, the proposed methodology could also be applied to such data, although there are better ways to study context effects using experiments, as the vast amount of prior research demonstrates. Appendix D presents an alternative application of the proposed methodology to the data

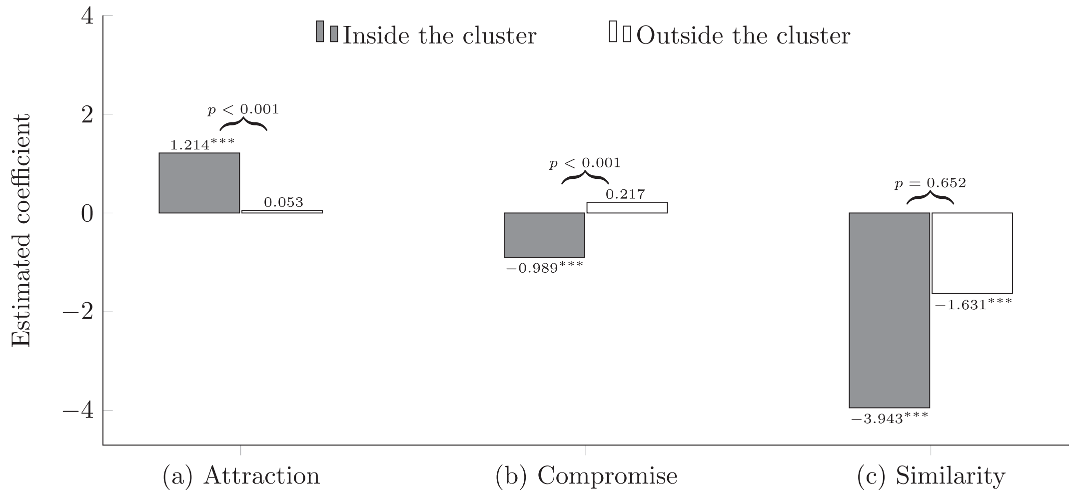

FIGURE 5 | Comparison of estimated coefficients inside and outside the cluster. Estimates derived from the model that augments model 6 by decomposing each measure across inside and outside the cluster using random effects probit estimation. p values are for the χ^2 tests for coefficient differences. The null hypothesis in each case is that the two coefficients are not different from each other.

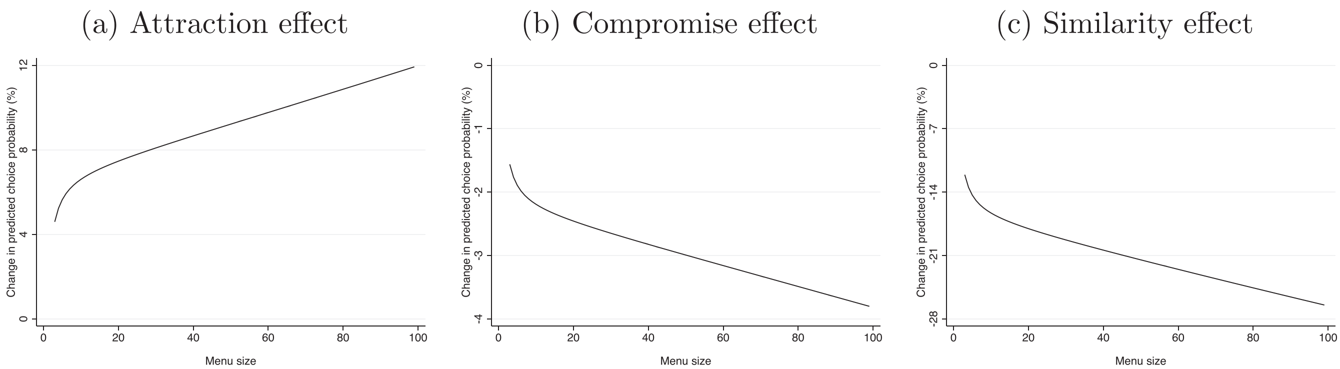

FIGURE 6 | Predicted effect on choice probability. The effect of increasing the corresponding measure from the mean by one standard deviation in our data sample. Effects calculated using model 6 while keeping all other variables at sample means.

from an experiment that was designed to study an unrelated question (Gerzinic et al. 2021). That lack of connection is important to interpreting experimental data as observational from our perspective. This experimental data allows us to overcome both drawbacks present in the current application. Additionally, the context of the experiment is transportation choice. The drawback therein is the fact that the experimental design by the authors does not introduce any variation in compromise measure. As a result, we are only able to study attraction and similarity effects. Here, we are guaranteed that users scrutinized all available options and have not only the identifier for a user, but also their characteristics that allow us to produce more reliable estimates. Again, our methodology does detect both attraction and similarity effects in directions predicted by the theory.

4.4 | Implications

The main implication of the paper is that context effects can be measured in less-sterile environments where consumers usually operate. In this specific case, context effects can be used by managers to optimize their revenue streams. Consider an airline operating a flight and being able to change its attributes. Even though the margins from Figure 3 are informative for understanding the change that an alteration in a given context measure can cause, they do not contextualize the magnitude of the change that managers can impose on those very attributes. Consider a manager that is supplying an average product on the market (a flight with all attributes and context measures set at our sample mean values) and is able to increase the context measures from their (sample) mean values by one (sample) standard deviation. This clearly changes the context in which consumers operate and affects choice probabilities. Figure 6 measures the effects in predicted choice probabilities across the three context measures for varying menu sizes. These results indicate that even though the magnitude of estimated margins is negatively associated with the menu size (see Figure 3), the association between the magnitude of the effect on choice probabilities and the menu size is positive for all three context effects. The reason for this is that margins decrease at a lower rate than the average choice probability of an option (which is equal to $1/M$). Hence, the overall impact that managers can have on the sales by manipulating the choice context is greater for larger menus.

One point worth noting here is that the above example considers changes in one context measure without altering the other two context measures. In fact, it also assumes that the attributes of the option are not altered. This is the idea of *ceteris paribus* analyses that we can read from statistical estimates. This, however, is impossible to reproduce in the field. All context measures depend on product attributes; indeed, they depend on the attributes of the product that a manager can control, as well as those of all other alternatives in the menu. Thus, changing the context measure of an option requires changing at least some attributes. Any such change would have a direct effect on choice probabilities, but it would also potentially affect some other context measures, which would have knock-on effects on the choice probabilities of the focal option and those of the competitors. In other words, taking into account the choice context makes the choice model complex, where a change in one attribute of one option would affect virtually all choice probabilities. In the Appendix F, we present a discussion of the example that takes estimates from the analysis in Section 4at face value and demonstrates how the overall effects of a specific option redesign can be calculated.

One important assumption that we have made to this point is that estimated context effects operate independently from one another. This is a common assumption (see, for example, Rooderkerk et al. 2011) that simplifies the estimation procedure and the interpretation of results. Research in psychology, however, has developed multiple models where one process can generate several context effects (see, e.g., Trueblood 2022). If multiple context effects are generated by one underlying choice process, it is possible that the context effects generated will be interdependent. Given the flexibility of our approach, we can study the interdependence between the context effects considered. Appendix G (Table G1) presents regression results with interactions between each pair of context effects. Out of three pairs, the only interaction that is precisely estimated (i.e., that reaches a reasonable level of statistical significance) is the one between attraction and compromise effects. The interaction coefficient is negative and highly significant, pointing to a strong interdependence between attraction and compromise. Including such interaction, however, renders the coefficient for the compromise measure positive and significant in the case of full dataset regressions. This implies that in the absence of dominance in a choice menu, the compromise effect is positive, as theoretically predicted. However, as dominance arises in a menu and causes an attraction effect, the compromise effect quickly turns negative. The higher the attraction effect is associated with the more negative the compromise effect. Conversely, the attraction effect is largest in menus with low compromise values. The size of the attraction effect is negatively associated with the value of compromise in a menu. These results suggest that while attraction and compromise effects might arise from related (if not identical) choice processes, the similarity effect arises due to a different mechanism that is not related to the one generating attraction and compromise effects. However, these are only indications that require further research and elaboration.

5 | Conclusion

In this paper, we have built upon established context measures to make contextual choice modeling applicable to data from a wide range of environments for the first time. Previous approaches to the study of context effects were limited to choices among a small number of options (usually two or three) of simple products (usually characterized by two numerical attributes). This meant much of the previous literature used experimental settings and examined relatively simple product line design questions (e.g., Orhun 2009; Rooderkerk et al. 2011), where the choice was among a set of similar products with marginally varied characteristics. Our framework opens up opportunities to study context effects in a wider range of settings. The methodology presented can handle large menus of relatively complex products characterized by many and varied types of attributes. Equally importantly, the approach does not require constancy of menu sizes across choice settings. In this paper we have presented one such empirical application to observational air travel choice data. However, this methodology could be applied to even larger and richer datasets generated from electronic commerce websites, as well as offline environments. This enables the study of context effects in the field without explicit experimental interventions, which are usually expensive. As a result, the methodology could contribute to designing optimal offers on a larger scale, especially in online settings where each outlet can carry hundreds of substitute options to choose from.

A significant drawback of our methodology is that for attraction and compromise effects it relies on ordinal relationships between available options. In other words, we use the information that option A is better than option B in terms of attribute X. This is a departure from previous research that has relied on cardinal measurements (e.g., Rooderkerk et al. 2011): that is, how much one option differs from another. When a researcher is dealing with two or three options, cardinal measurement is necessary to quantify context. In our proposed framework, on the other hand, cardinal measurement is not a necessity. However, given that cardinal measurement is possible in most settings (at least for some attributes), including both cardinal and ordinal measures (which capture more substantial component of variance) could potentially allow for fine-tuning the analysis. We leave the development of context measures that will consider both ordinal and cardinal measures to future work. By developing novel measures of context effects that can be applied to large-scale observational datasets with a varying number of alternatives in the menu, this paper opens up a new direction toward understanding how certain aspects of behavioral sciences can be studied in the context of large-scale observational real-world data.

Ethics Statement

This article does not contain any studies with human participants or animals performed by any of the authors. As a result, the requirement of IRB approval was waived.

Consent

For this type of study, informed consent is not required.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Endnotes

[1] A notable exception is a recent study by Stanley and Wedell (2024), who examine the presence of certain context effects in menus of up to 15 options.

[2] Recent examples of this approach are linear context logit by Tomlinson and Benson (2021) and contextual multinomial logit by Maragheh et al. (2020).

[3] Interestingly, the centroid model of Bodner and Prelec (1994) can also account for the attraction, but not for similarity effect. A more flexible accummulation-to-threshold class of models can in fact account for all three context effects simultanously (Roe et al. 2001).

[4] Combining frequency and distance measures in one metric requires arbitrage across the two drivers of context effects. It is not clear how to solve such a problem (i.e., it is not clear if greater attraction effect is generated by two decoys targeting the same alternative or by a unique decoy located twice further from the target). Evidence in existing empirical research highlights the complexity of the problem. While Castillo (2020) reports that doubling the distance between the decoy and the target increases the attraction effect, Daviet and Webb (2023) show that introducing the second decoy without changing the range of the attributes also increases the attraction effect.

[5] An alternative measure of attraction could compute the difference between the number of options that the focal alternative dominates and the number of options by which the focal alternative is dominated. This procedure produces a measure that is confined to $(-1;1)$ instead of the one confined to $(0;1)$. This generates higher variance and accentuates differences between alternatives at the lower end of the distribution (i.e., those that are largely dominated by other options in the menu). Given that the alternative measure does not significantly alter the right tail of the attraction distribution (which constitutes our point of interest), we prioritize the simplicity of calculation and interpretation and opt for the definition given by Equation (3).

[6] An alternative way to quantify the compromise between two sets of extreme options is to count the number of all possible pairs for which a given focal option is a compromise. This would result in $Comp'(F) = O(Group_3) * O(Group_4)$. This measure behaves very similarly to the one discussed in the paper. In fact, the correlation between the two compromise measures in the dataset that we use in this paper is 0.825. All results reported in the paper are qualitatively unaltered by the replacement of the compromise measure with this alternative. However, we prefer working with the compromise measure in the paper as it takes a more comprehensive view of the choice process.

[7] A straightforward alternative to such a definition would be to use a different value of r. However, there is no valid guiding principle that could help us choose the optimal cutoff value. As a result, in the empirical application in Section 4, we test the stability of our results by re-running the analysis with an alternative value of $r = 0.1$. Another alternative could be to use the clustering method to identify options similar to a given focal alternative and use the number of such alternatives as a similarity measure. Such an alternative measure is highly correlated with the measure we use in this paper (0.797, when using affinity propagation clustering method) and does not qualitatively alter the analysis in Section 4.2. However, this alternative measure is computationally more demanding and its usage would have precluded us from producing the analysis in Section 4.3 to alleviate the concerns presented by the specific nature of our empirical application.

[8] Office ID, trip origin and destination, trip dates and number of passengers are matched exactly. The distance in time between a booking and the preceding search is minimized. If, given exactly matched attributes, the booking was not performed within 24 h after a given search, the search is deemed unmatched. If, given exactly matched attributes, we find no search during 24 h preceding a given booking, the booking is deemed unmatched. All unmatched searches and bookings are dropped from the analysis.

[9] In rare cases the route between two European airports could go through a different continent.

[10] Note that Table 1 reports the mean value of menu size 73.91, while previously we have stated that the average menu size is 60. The reason for this discrepancy is that Table 1 presents option-level descriptives, while the average menu size is a menu-level measure.

[11] The results from the sensitivity analysis to this cutoff value (i.e., recalculation of the context using a stricter similarity definition of $r = 0.1$) is presented in Appendix B.

[12] The classical procedures for fixed effects probit model fail to converge in our data. We overcome the problem by manually removing menu-level averages from all vertical variables and fitting the plain probit morel. This estimation gives comparable estimates to those obtained from random effects probit (see Table C1 in Appendix C).

[13] For the regression analysis, similar to other numeric variables, in order to eliminate any scale effects, we perform a z-score transformation of the duration of stay variable.

[14] We have also explored the behavior of the alternative attraction measure presented in footnote 5. The correlation between the two attraction measures is 0.89 and the results are consistent across estimations of models 2 and 6.

[15] Along with parameter estimates and standard errors, Table 2 reports in-sample model fit metrics. We do not study the out-of sample predictive power of our models because our aim is to study the potential presence of context effects in consumer choice, rather than the identification of the best performing model in order to predict consumer choice. The latter will clearly depend on the specific empirical application of the proposed methodology.

[16] Affinity propagation algorithm, however, does need prespecification of one parameter—the damping factor (λ) (see Appendix D). In what follows, we present the results for $\lambda = 0.5$, which has been a standard value in the literature (Frey and Dueck 2007; Bauer and Schedl 2019).

[17] The magnitude (margin) of the similarity effect does not change much as we move from complete menus to the reduced dataset. This is due to the fact that clustering algorithms overlap significantly with our similarity measure. Clustering eliminates distant options reducing the range across each dimension and squeezing the set of alternatives that were regarded "similar" to the focal option in the original estimation.

[18] We need a minimum of two dimensions that can be compared across two comparable groups.

[19] It has the second-closest values in both price and flight duration to the menu mean values of those two attributes. Note that the closest options within each attribute are ranked very low in the other attribute.

[20] To keep the simulated scenario closer to reality, in both cases reduction in the value of one attribute is achieved by an increase in the value of another attribute. However, this simulation can be extended to any sort of redesign of the menu.

References

Adler, S. J., M. K. Schöniger, M. Lichters, and M. Sarstedt. 2024. "Forty Years of Context Effect Research in Marketing: A Bibliometric Analysis." *Journal of Business Economics* 94: 437–466.

Apffelstaedt, A., and L. Mechtenberg. 2021. "Competition for Context-Sensitive Consumers." *Management Science* 67, no. 5: 2828–2844.

Bauer, C., and M. Schedl. 2019. "Global and Country-Specific Mainstreaminess Measures: Definitions, Analysis, and Usage for Improving Personalized Music Recommendation Systems." *PLoS ONE* 14, no. 6: 1–36.

Ben-Akiva, M. E., and S. R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Predict Travel Demand*. Cambridge, MA: MIT Press.

Bergner, A. S., D. M. Oppenheimer, and G. Detre. 2019. "Vamp (Voting Agent Model of Preferences): A Computational Model of Individual Multi-Attribute Choice." *Cognition* 192: 103971.

Bhatia, S. 2013. "Associations and the Accumulation of Preference." *Psychological Review* 120, no. 3: 522–543.

Bodner, R. and D. Prelec (1994). The Centroid Model of Context-Dependent Choice. Working Paper, Sloan School of Management.

Bower, A. and L. Balzano (2020). Preference Modeling With Context-Dependent Salient Features. In *Proceedings of the 37th International Conference on International Conference on Machine Learning*, Volume 119, pp. 1067–1077.

Castillo, G. 2020. "The Attraction Effect and Its Explanations." *Games and Economic Behavior* 119: 123–147.

Cataldo, A. M., and A. L. Cohen. 2019. "The Comparison Process as an Account of Variation in the Attraction, Compromise, and Similarity Effects." *Psychonomic Bulletin & Review* 26: 934–942.

Chen, S. and T. Joachims (2016). Predicting Matchups and Preferences in Context. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Volume 22, pp. 775–784.

Chernev, A. 2004. "Extremeness Aversion and Attribute-Balance Effects in Choice." *Journal of Consumer Research* 31, no. 2: 249–263.

Daviet, R., and R. Webb. 2023. "A Test of Attribute Normalization via a Double Decoy Effect." *Journal of Mathematical Psychology* 113: 102741.

Dhar, R., S. M. Nowlis, and S. J. Sherman. 2000. "Trying Hard or Hardly Trying: An Analysis of Context Effects in Choice." *Journal of Consumer Psychology* 9, no. 4: 189–200.

Dotson, J. P., J. R. Howell, J. D. Brazell, et al. 2018. "A Probit Model With Structured Covariance for Similarity Effects and Source of Volume Calculations." *Journal of Marketing Research* 55, no. 1: 35–47.

Dowling, K., D. Guhl, D. Klapper, M. Spann, L. Stich, and N. Yegoryan. 2020. "Behavioral Biases in Marketing." *Journal of the Academy of Marketing Science* 48, no. 3: 449–477.

Doyle, J. R., D. J. O'Connor, G. M. Reynolds, and P. A. Bottomley. 1999. "The Robustness of the Asymmetrically Dominated Effect: Buying Frames, Phantom Alternatives, and In-Store Purchases." *Psychology & Marketing* 16, no. 3: 225–243.

Drolet, A. 2002. "Inherent Rule Variability in Consumer Choice: Changing Rules for Change's Sake." *Journal of Consumer Research* 29, no. 3: 293–305.

Evangelidis, I., S. Bhatia, J. Levav, and I. Simonson. 2024. "50 Years of Context Effects: Merging the Behavioral and Quantitative Perspectives." *Journal of Consumer Research* 51, no. 1: 19–28.

Evangelidis, I., J. Levav, and I. Simonson. 2018. "The Asymmetric Impact of Context on Advantaged Versus Disadvantaged Options." *Journal of Marketing Research* 55, no. 2: 239–253.

Frederick, S., L. Lee, and E. Baskin. 2014. "The Limits of Attraction." *Journal of Marketing Research* 51, no. 4: 487–507.

Frey, B. J., and D. Dueck. 2007. "Clustering by Passing Messages Between Data Points." *Science* 315, no. 5814: 972–976.

Fridman, A., O. Amir, and K. T. Hansen. 2024. "Dominance Effects in the Wild." *Journal of Consumer Research* 50: 1117–1135.

Gerzinic, N., S. van Cranenburgh, O. Cats, E. Lancsar, and C. Chorus. 2021. "Estimating Decision Rule Differences Between "Best" and "Worst" Choices in a Sequential Best Worst Discrete Choice Experiment." *Journal of Choice Modelling* 41: 100307.

Gilbride, T. J., and G. M. Allenby. 2004. "A Choice Model With Conjunctive, Disjunctive, and Compensatory Screening Rules." *Marketing Science* 23, no. 3: 391–406.

Guo, L. 2022. "Testing the Role of Contextual Deliberation in the Compromise Effect." *Management Science* 68, no. 6: 4326–4355.

He, J. 2024. "Bayesian Contextual Choices Under Imperfect Perception of Attributes." *Management Science* 70, no. 3: 1465–1482.

Heller, D., I. P. Levin, and M. Goransson. 2002. "Selection of Strategies for Narrowing Choice Options: Antecedents and Consequences." *Organizational Behavior and Human Decision Processes* 89, no. 2: 1194–1213.

Huber, J., J. W. Payne, and C. Puto. 1982. "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis." *Journal of Consumer Research* 9, no. 1: 90–98.

Huber, J., and C. Puto. 1983. "Market Boundaries and Product CHOICE: Illustrating Attraction and Substitution Effects." *Journal of Consumer Research* 10, no. 1: 31–44.

Kahneman, D., and A. Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47, no. 2: 263–291.

Kamakura, W. A., and R. K. Srivastava. 1984. "Predicting Choice Shares Under Conditions of Brand Interdependence." *Journal of Marketing Research* 21, no. 4: 420–434.

Kivetz, R., O. Netzer, and V. Srinivasan. 2004. "Alternative Models for Capturing the Compromise Effect." *Journal of Marketing Research* 41, no. 3: 237–257.

Köcher, S., M. Jugovac, D. Jannach, and H. H. Holzmüller. 2019. "New Hidden Persuaders: An Investigation of Attribute-Level Anchoring Effects of Product Recommendations." *Journal of Retailing* 95, no. 1: 24–41.

Kumar Padamwar, P., V. Kumar Kalakbandi, and J. Dawra. 2023. "Deliberation Does Not Make the Attraction Effect Disappear: The Role of Induced Cognitive Reflection." *Journal of Business Research* 154: 113335.

Lehmann, D. R., and Y. Pan. 1994. "Context Effects, New Brand Entry, and Consideration Sets." *Journal of Marketing Research* 31, no. 3: 364–374.

Lichters, M., M. Sarstedt, and B. Vogt. 2015. "On the Practical Relevance of the Attraction Effect: A Cautionary Note and Guidelines for Context Effect Experiments." *AMS Review* 5, no. 1: 1–19.

Lloyd, S. 1982. "Least Squares Quantization in PCM." *IEEE Transactions on Information Theory* 28, no. 2: 129–137.

Luce, R. D. 1959. *Individual Choice Behavior: A Theoretical Analysis*. New York, NY: John Wiley & Sons.

Mao, W. 2016. "When One Desires Too Much of a Good Thing: The Compromise Effect Under Maximizing Tendencies." *Journal of Consumer Psychology* 26, no. 1: 66–80.

Maragheh, R. Y., X. Chen, J. Davis, J. Cho, and S. Kumar (2020). "Choice Modeling and Assortment Optimization in the Presence of Context Effects". SSRN preprint: https://doi.org/10.2139/ssrn.3747354.

Marini, M., A. Sapienza, and F. Paglieri. 2023. "There Is More to Attraction Than Meets the Eye: Studying Decoy-Induced Attention Allocation Without Eye Tracking." *Journal of Behavioral Decision Making* 36, no. 2: e2299.

McFadden, D. 2001. "Economic Choices." *American Economic Review* 91, no. 3: 351–378.

Mirzayev, E., Z. Babutsidze, W. Rand, and T. Delahaye (2021). "Use of Clustering for Consideration Set Modelling in Recommender Systems". In *Proceedings of the 54th Hawaii International Conference on System Sciences*, Volume 54, pp. 4270–4278.

Moe, W. W. 2006. “An Empirical Two-Stage Choice Model With Varying Decision Rules Applied to Internet Clickstream Data.” *Journal of Marketing Research* 43, no. 4: 680–692.

Molter, F., A. W. Thomas, S. A. Huettel, H. R. Heekeren, and P. N. C. Mohr. 2022. “Gaze-Dependent Evidence Accumulation Predicts Multi-Alternative Risky Choice Behaviour.” *PLoS Computational Biology* 18, no. 7: e1010283.

Mottini, A. and R. Acuna-Agost (2017). Deep Choice Model Using Pointer Networks for Airline Itinerary Prediction. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Volume 23, pp. 1575–1583.

Mousavi, N., P. Adamopoulos, and J. Bockstedt. 2023. “The Decoy Effect and Recommendation Systems.” *Information Systems Research* 34, no. 4: 1533–1553.

Noguchi, T., and N. Stewart. 2014. “In the Attraction, Compromise, and Similarity Effects, Alternatives Are Repeatedly Compared in Pairs on Single Dimensions.” *Cognition* 132, no. 1: 44–56.

Orhun, A. Y. 2009. “Optimal Product Line Design When Consumers Exhibit Choice Set-Dependent Preferences.” *Marketing Science* 28, no. 5: 868–886.

Otto, A. R., S. Devine, E. Schulz, A. M. Bornstein, and K. Louie. 2022. “Context-Dependent Choice and Evaluation in Real-World Consumer Behavior.” *Scientific Reports* 12, no. 1: 17744.

Pettibone, J. C., and D. H. Wedell. 2000. “Examining Models of Nondominated Decoy Effects Across Judgment and Choice.” *Organizational Behavior and Human Decision Processes* 81, no. 2: 300–328.

Pfannschmidt, K., P. Gupta, B. Haddenhorst, and E. Hüllermeier. 2022. “Learning Context-Dependent Choice Functions.” *International Journal of Approximate Reasoning* 140: 116–155.

Pinger, P., I. Ruhmer-Krell, and H. Schumacher. 2016. “The Compromise Effect in Action: Lessons From a Restaurant’s Menu.” *Journal of Economic Behavior and Organization* 128: 14–34.

Rafaï, I., Z. Babutsidze, T. Delahaye, N. Hanaki, and R. Acuna-Agost. 2022. “No Evidence of Attraction Effect Among Recommended Options: A Large-Scale Field Experiment on an Online Flight Aggregator.” *Decision Support Systems* 153: 113672.

Roe, R. M., J. R. Busemeyer, and J. T. Townsend. 2001. “Multialternative Decision Field Theory: A Dynamic Connectionist Model of Decision Making.” *Psychological Review* 108, no. 2: 370–392.

Rokach, L., and O. Maimon. 2005. *Clustering Methods*, 321–352. Boston, MA: Springer.

Rooderkerk, R. P., H. J. V. Heerde, and T. H. Bijmolt. 2011. “Incorporating Context Effects Into a Choice Model.” *Journal of Marketing Research* 48, no. 4: 767–780.

Simonson, I. 1989. “Choice Based on Reasons: The Case of Attraction and Compromise Effects.” *Journal of Consumer Research* 16, no. 2: 158–174.

Spektor, M. S., D. Kellen, and K. C. Klauer. 2022. “The Repulsion Effect in Preferential Choice and its relation to Perceptual Choice.” *Cognition* 225: 105164.

Stanley, J. M., and D. H. Wedell. 2024. “Impact of Choice Set Complexity on Decoy Effects.” *Journal of Behavioral Decision Making* 37, no. 2: e2373.

Tomlinson, K. and A. R. Benson (2021). Learning Interpretable Feature Context Effects in Discrete Choice. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, Volume 27, pp. 1582–1592.

Trueblood, J. S. 2022. “Theories of Context Effects in Multialternative, Multiattribute Choice.” *Current Directions in Psychological Science* 31, no. 5: 428–435.

Trueblood, J. S., S. D. Brown, A. Heathcote, and J. R. Busemeyer. 2013. “Not Just for Consumers: Context Effects Are Fundamental to Decision Making.” *Psychological Science* 24, no. 6: 901–908 PMID: 23610134.

Tversky, A. 1972. “Elimination by Aspects: A Theory of Choice.” *Psychological Review* 79, no. 4: 281–299.

Tversky, A., and I. Simonson. 1993. “Context-Dependent Preferences.” *Management Science* 39, no. 10: 1179–1189.

Wedell, D. H., and J. C. Pettibone. 1996. “Using Judgments to Understand Decoy Effects in Choice.” *Organizational Behavior and Human Decision Processes* 67, no. 3: 326–344.

Wollschlaeger, L. M., and A. Diederich. 2020. “Similarity, Attraction, and Compromise Effects: Original Findings, Recent Empirical Observations, and Computational Cognitive Process Models.” *American Journal of Psychology* 133, no. 1: 1–30.

Wu, C., and K. Cosguner. 2020. “Profiting From the Decoy Effect: A Case Study of an Online Diamond Retailer.” *Marketing Science* 39, no. 5: 974–995.

Appendix A

Comparison Pair Count for Compromise Effect

Given a setting with N vertical attributes, there are a number of ways a focal option can act as a compromise between two groups of competing options. We discuss this case by case.

Case 1: No dimension is equal across focal and competing options. In this case one group of options might be better than the focal option in one dimension and worse in $N-1$ dimensions. The mirror image of this group would be a group that is better than the focal option in $N-1$ dimensions and worse in the same one dimension. There are $\binom{N}{1}$ of such groups.
Another option within the same case could be a group that is better than focal option in two dimensions and worse in $N-2$ dimensions. There are $\binom{N}{2}$ such groups.
All in all, there are $N-1$ such sub-cases. These sub-cases count as distinct groups. As we have to pair these groups, we divide the number by two.
Therefore, for this case the number of comparisons is $\frac{1}{2}\sum_{a=1}^{N-1}\binom{N}{a}$.

Case 2: Only one dimension is equal across all competing options, including the focal option. In this case, we are down to comparing not N, but $N-1$ dimensions. Therefore, for every given dimension that is equal across options, we have $\frac{1}{2}\sum_{a=1}^{N-2}\binom{N-1}{a}$ comparisons. However, not only one but each of the N vertical attributes can be equal across all options. Therefore, the total number of comparisons in this case is $\frac{1}{2}\binom{N}{1}\sum_{a=1}^{N-2}\binom{N-1}{a}$.

Case 3: Multiple dimensions are equal across all competing options. First, we extend the previous case to the situation where two dimensions are equal across all options, which yields $\frac{1}{2}\binom{N}{2}\sum_{a=1}^{N-3}\binom{N-2}{a}$. We iterate the same exercise until (and including) the setup where we have $N-2$ dimensions that are equal across all options.[18]

The total of all comparisons will simply be the sum of all these cases, which can be expressed as

$$\Omega = \frac{1}{2}\sum_{b=0}^{N-2}\left[\binom{N}{b}\sum_{a=1}^{N-1-b}\binom{N-b}{a}\right] \tag{A1}$$

Appendix B

Sensitivity Check of the Threshold (r) for the Similarity Measure

All results in the body of the paper are presented with the threshold value of $r = 0.2$ for the similarity measure. Table B1 presents results for a stricter definition of similarity—$r = 0.1$—for models that involve similarity (models 5 and 6). All results are qualitatively comparable to those presented in Table 2.

TABLE B1 | Estimation results with $r = 0.1$ as a threshold level for similarity.

	Model 5	**Model 6**
Price	−0.243***	−0.225***
	(0.007)	(0.009)
Trip duration	−0.112***	−0.088***
	(0.007)	(0.011)
Number of flights	−0.315***	−0.314***
	(0.007)	(0.007)
Number of airlines	−0.208***	−0.199***
	(0.009)	(0.009)
Menu size	−0.016***	−0.016***
	(0.000)	(0.000)
Attraction		0.119*
		(0.042)
Compromise		−0.288*
		(0.094)
Similarity	−3.114***	−3.157***
	(0.519)	(0.522)
Constant included	Yes	Yes
Horizontal variables as controls	Yes	Yes
Number of observations	344,386	344,386
Number of individual	5784	5784
Akaike information criterion	45,196	45,182
Log likelihood	−22,582	−22,572

Note: Results from random effects probit estimations. Complete estimation output can be found in the Appendix I (Table I5). Groups represent menus/clusters in the datasets. Robust (menu-level clustered) standard errors are in parentheses. Significance key: "***" indicates $p < 0.001$, "**" indicates $p < 0.01$, "*" indicates $p < 0.05$.

Appendix C

Results From the Fixed Effects Estimation

As a sensitivity check for the random effects probit regression, this section presents results from the menu-level fixed effects probit regression. The fixed effects probit procedure on our data could not converge (due to computational complexity). Therefore, we have de-meaned all explanatory variables (and covariates) at the menu level and fitted a standard probit regression. A significant downside of the fixed effects approach is that we cannot control for the menu size because it is invariant at the menu level. Results are given in Table C1 and demonstrate that qualitative results are not sensitive to this alteration.

TABLE C1 | Estimation results with fixed effects.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Price	−0.237***	−0.245***	−0.204***	−0.240***	−0.240***	−0.194***
	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)
Trip duration	−0.092***	−0.109***	−0.047***	−0.109***	−0.108***	−0.046***
	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)
Number of flights	−0.364***	−0.320***	−0.321***	−0.313***	−0.306***	−0.302***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Number of airlines	−0.209***	−0.209***	−0.178***	−0.211***	−0.208***	−0.179***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Menu size	−0.015***	−0.016***	−0.016***	−0.016***	−0.016***	−0.017***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Attraction			0.005***			0.005***
			(0.001)			(0.001)
Compromise				−0.015***		−0.014***
				(0.002)		(0.002)
Similarity					−0.020***	−0.018***
					(0.003)	(0.003)
Constant included	Yes	Yes	Yes	Yes	Yes	Yes
Horizontal variables as controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	344,386	344,386	344,386	344,386	344,386	344,386
Number of choices	5784	5784	5784	5784	5784	5784
Akaike information criterion	46,315	45,454	45,438	45,445	45,303	45,277
Log likelihood	−23,151	−22,712	−22,703	−22,707	−22,635	−22,620

Note: Results from fixed effects probit estimations. Fixed effects estimation was achieved by de-meaning all variables at the menu level and then fitting a standard Probit procedure. Complete estimation output can be found in the Appendix I(Table I5). Standard errors are in parentheses. Significance key: "***" indicates $p < 0.001$, "**" indicates $p < 0.01$, "*" indicates $p < 0.05$.

Appendix D

Clustering

The goal of clustering is to separate data into different groups such that similar instances belong to the same group, while dissimilar instances are allocated to different groups (Rokach and Maimon 2005). Formally, clustering consists in making a partition $C = \{C_1, C_2, \cdots, C_k\}$ of some set S in a way that $S = \cup_{I=1}^{k} C_i$ and $C_i \cap C_j \neq 0 \, \forall \, i \neq j$. Thus, any alternative in set S belongs to exactly one cluster.

There are many clustering methods, each of which has a different way of defining similar items and as a result will group in different ways. Clustering is usually an unsupervised machine learning method in that there are no preconceived labels given to the clusters. This implies that we have no universal way of evaluating the quality of a clustering outcome. Affinity propagation (Frey and Dueck 2007) detects a limited number of "exemplars," which are identified as the best "representative" of other objects in the same cluster ("samples"). It calculates the pairwise values characterizing the suitability for one object to be the exemplar of the other. These values are updated in response to values from other pairs. This updating happens in an iterative manner until convergence, at which point the final exemplars are chosen, and hence the final clustering is identified.

There are two characteristics involved in the process. The responsibility $r(i,k)$ which quantifies how suitable k is as an exemplar of the cluster i compared to all other potential exemplars. It is calculated as

$$r(i,k) = s(i,k) - max[a(i,k') + s(i,k') \forall k' \neq k] \tag{D1}$$

where $s(i,k)$ is the similarity between i and k, measured as the negative squared error.

The second property is the availability $a(i,k)$, which measures the extent to which i is an appropriate sample of k, given all other already identified samples of k. This is calculated as

$$a(i,k) = min\left[0, r(k,k) + \sum_{i' \sim s.t. \sim i' \notin \{i,k\}} r(i',k)\right] \tag{D2}$$

At the start, both r and a are set to zero, and the calculations are iterated until full convergence. To eliminate oscillations when updating the values, the damping factor λ is introduced to the iteration process. This facilitates the convergence process and alters responsibility and availability equations as follows:

$$r_{t+1}(i,k) = \lambda r_t(i,k) + (1-\lambda) r_{t+1}(i,k) \tag{D3}$$

$$a_{t+1}(i,k) = \lambda a_t(i,k) + (1-\lambda) a_{t+1}(i,k) \tag{D4}$$

The damping factor, $\lambda \in [0;1]$, affects the number of identified clusters in the affinity propagation procedure. After setting its value, the number of clusters in the data are automatically identified. Frey and Dueck (2007) recommend setting $\lambda \in [0.5;1]$ in order to ensure convergence in large datasets. We have experimented with the sensitivity of clustering outcomes with respect to the damping factor and have found very little difference in the vicinity of the factor between 0.5 and 0.75.

Appendix E

Application to Experimental Data

Air traveler choice data presents an excellent opportunity for applying the proposed methodology. It is a large dataset of actual choices made in a natural environment by consumers, the product is relatively complex in that it is characterized by more than two attributes, the menu size is not constant across different choice cases, and menu sizes are sufficiently large to obtain variance in all context variables. However, the dataset also has shortcomings. First, even though we know what was available on the market when the choice was made, we do not have accurate information on which options were considered by the consumer. Second, we do not have information on the identity and characteristics of the consumers. Without such information, we are not able to account for consumer-side features that could systematically drive the choice outcomes that we observe.

To remedy these shortcomings, in what follows, we apply the same methodology to an experimental dataset. This data, like the observational dataset, comes from a travel context. Similarly, the studied product is relatively complex. Unlike the air travel observational dataset, however, this dataset comes from a stated choice experiment. Here, we do not have variance in menu sizes, and these menus are relatively small, with five alternatives. Importantly, this dataset has information on a set of variables describing decision maker demographics. An added advantage of the dataset is that each subject is making 12 choices and that these 12 choice scenarios are constant across all subjects. This allows us to control for menu-specific and subject-specific characteristics.

The commuter choice data comes from a discrete choice experiment administered to residents of and daily commuters to the city of Ljubljana, Slovenia, by Gerzinic et al. (2021). A total of 108 subjects were sequentially presented with 12 five-alternative menus and were asked to choose the best alternative in each case, leading to 1296 recorded choice cases. Each alternative described a commuter trip with a "park and ride facility choice" to the city with respect to the following five characteristics: price, car ride duration, public transport ride duration, average public transport wait time and the mode of public transport (either bus or train). Subjects were mainly Slovenian nationals (91.67%), 58% female and had a mean age of 36 years (St. Dev. = 12.5). Further information on education, income, household size and the number of cars in that household was also obtained. Descriptive statistics of the choice variables in the experimental dataset are presented in Table E1. Subject characteristics are used as control variables.

TABLE E1 | Descriptive statistics for choice and context variables in experimental commuter choice dataset.

Variable	# of obs.	Mean	S.D.	Min.	Max.
Price	6480	5	3.266	1	9
Car ride duration	6480	15	8.166	5	25
Public transport ride duration	6480	20	8.166	10	30
Public transport wait time	6480	16.67	10.275	5	30
Public transport is train (dichotomous variable)	6480	0.50	0.500	0	1
Attraction	6480	0.067	0.249	0	1
Similarity	6480	2.633	1.08	1	4

Measurement of Context

A significant disadvantage of this particular dataset is its small menu size (five) and large number of choice variables (also five). These circumstances, together with the fact that the experiment was not designed for the specific purpose of studying context effects and that one of the choice variables (mode of public transport) is categorical, restricts the variance in context variables. As a result, our procedure does not identify a single case in which we can observe a compromise option, which makes it impossible to study. In addition, the presence of the categorical variable drives the affinity propagation clustering algorithm (and the K-means algorithm, for that matter), which always results in two groups of similar options—one consisting of all options using the bus as the mean of public transport, and the other using the train. Given that calculating the dominance relationship also requires a constant categorical variable across a pair of options, all dominance relationships (and thus all attraction effects) are only within the cluster. The consequence of all of the above is that we cannot estimate model 4, and that models 6 through 9 all become equivalent. Table E1 also presents descriptive statistics of context variables.

Choice Modeling

This choice modeling exercise takes a very similar approach to the one using observational data. The only difference is that the current case regressions also include menu-level fixed effects. This is necessary as 12 choice cases are constant for all subjects. The results of the random effects probit estimation are given in Table E2. At the choice variable level, we find negative effects of all vertical variables (price, car ride duration, public transport ride duration, and public transport wait time). We also find that commuters prefer the train over the bus as a mode of public transportation.

The results with respect to context variables are consistent with those obtained from the observational dataset in that we find significant attraction and similarity effects. Even though the attraction effect in model 3 is insignificant and has a negative sign, in the unifying model of context effects it achieves statistical significance (p-value = 0.021). Marginal effects indicate a much stronger impact of the context in the experimental setup compared to the similar exercise with observational data. A one-unit increase in attraction results in an increase of 4.8 percentage points in the choice likelihood of an option. A one-unit increase in similarity, on the other hand, decreases the choice likelihood by 3.2 percentage pints, *ceteris paribus*. This difference in marginal effects is not surprising given the much smaller menu size compared to the observational dataset.

TABLE E2 | Estimation results.

	Model 1	Model 2	Model 3	Model 4	Model 5
Price	−0.186***	−0.191***	−0.190***	−0.183***	−0.177***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Car ride duration	−0.063***	−0.066***	−0.063***	−0.065***	−0.059***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Public transport ride duration	−0.038***	−0.038***	−0.037***	−0.034***	−0.031***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Public transport wait time	−0.036***	−0.037***	−0.037***	−0.033***	−0.030***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Public transport is train	0.162***	0.153**	0.153**	0.130**	0.119*
	(0.047)	(0.049)	(0.049)	(0.049)	(0.050)
Attraction			1.289***		2.220***
			(0.348)		(0.361)
Similarity				−0.471***	−0.699***
				(0.103)	(0.109)
Constant included	Yes	Yes	Yes	Yes	Yes
Control variables included	No	Yes	Yes	Yes	Yes
Subject-level fixed effects included	Yes	Yes	Yes	Yes	Yes
Number of observations	6480	6180	6180	6180	6180
Number of choices	1296	1236	1236	1236	1236
Akaike information criterion	5455	5143	5132	5123	5093
Log likelihood	−2613	−2462	−2456	−2451	−2436

Note: Results from random effects probit estimations. Groups represent menus in the datasets. Robust (menu-level clustered) standard errors are in parentheses. Control variables include gender, age and education level of the experimental subject, as well as income level, number of cars and size of household. Significance key: "***" indicates $p < 0.001$, "**" indicates $p < 0.01$, "*" indicates $p < 0.05$.

Appendix F

An Example of Calculating the Overall Impact of Product Redesign

The estimates from section 4 can be used to calculate the overall effect of product alteration (i.e., changing one or more attributes of one or more alternative in a menu) while taking into account all measured context effects. Consider the 40-option menu displayed in Figure F1. Note that consumers prefer alternatives that have lower values across each of the two displayed characteristics. This example menu was generated by randomly drawing price and flight duration numbers from uniform distributions. We consider that all these options are exactly equal in all other attributes and set all those values at sample means from our dataset. Using our model estimates from Table 2, we can calculate predicted choice probabilities. Note that this exercise requires an additional layer of normalizing probabilities implied by the model so that they add up to one. Consider an option A (with price of 253 and flight duration 222) in the menu. This is an option that can be considered the most "average" in the menu.[19] Not surprisingly, according to model 2 (the model without context measures), this option is indeed most average among all alternatives – its probability of 2.4% is closest to the overall average (2. 5%). Model 2, however, does not take choice context into account. Measuring the context indicates that option A has a value of 0.275 for attraction, 0.527 for compromise and 0.075 for similarity. Being able to take the choice context into account (based on three measures of this option, as well as measures of all other options), model 6 assigns a much lower choice probability of 1.5% to option A.

Let us now consider a manager who is in control of redesigning product features and considers two exercises: replacing option A with option A' or replacing it with option A''.[20] Moving an option from the A to the A' location reduces the choice probability predicted by model 2 from 2.4% to 1.9%. Such a move, however, also changes the choice context. As far as context measures of the focal product are concerned, attraction and similarity values stay the same, while the compromise measure decreases from 0.527 to 0.095. This overcompensates the drop predicted by model 2; model 6 has a qualitatively different prediction that the move will increase the choice probability (from 1.5% to 1.7%).

Being able to alter only one context measure by changing attributes (as in the example above) of an option is something of an exception. Most of the product redesigns would change all three context measures of the focal option at once, along with those of many other alternatives in the menu. Moving an option from location A to location A'' does just that. It decreases all three context measures for the focal option to 0.150, 0.232 and 0.05 for attraction, compromise and similarity respectively. Such a redesign is again evaluated differently by our models with and without context effects. While model 2 predicts a drop in choice probability (from 2.4% to 2.3%), model 6 predicts a sizable increase (from 1.5% to 2.2%).

Again, due to the fact that the context-dependent choice model is complex, altering context measures of one option has a much more significant effect on the choice probabilities of other options in the menu than the model that does not take choice context into account. In the latter case, this effect is strictly driven by renormalization of choice probabilities. In the former case, the main driving force is changing context. Figure F2 plots kernel densities of changes in probabilities associated with the remaining 39 alternatives in the menu as a result of the option A redesign. These are very limited in case of model 2, but are clearly more consequential for model 6. Managers need to take these complexities into account when designing their contributions to the overall menu that is being offered to a consumer.

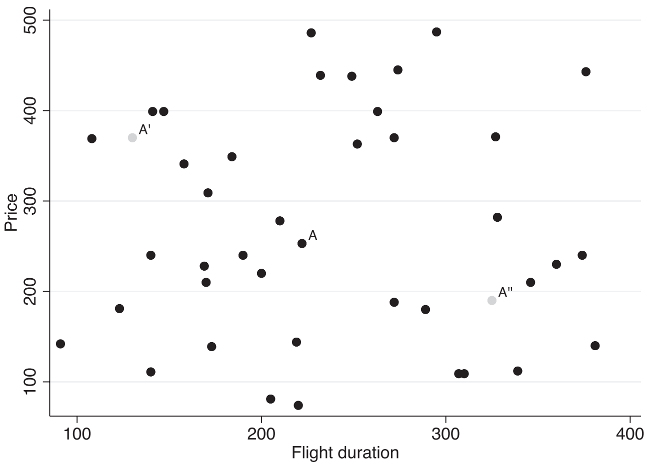

FIGURE F1 | Changing attributes of an option (A).

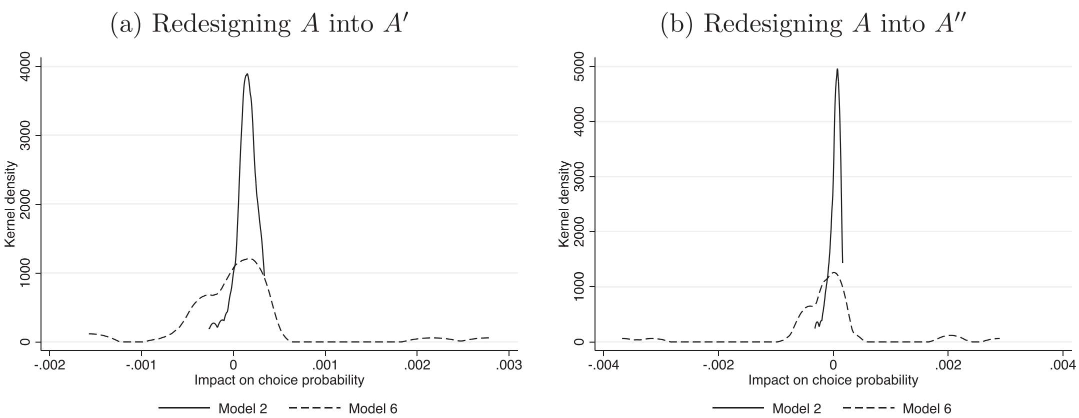

FIGURE F2 | Probability changes associated to product redesign. *Note:* Estimated kernel densities for probability changes as a result of redesigning option A for remaining 39 options in the menu.

Appendix G

Estimations With Context Effect Interactions

TABLE G1 | Regression outputs from regressions with interactions between pairs of context effects.

	Full dataset			Reduced dataset		
	Model 7	**Model 8**	**Model 9**	**Model 7**	**Model 8**	**Model 9**
Price	−0.217***	−0.219***	−0.218***	−0.090***	−0.090***	−0.091***
	(0.009)	(0.009)	(0.009)	(0.005)	(0.005)	(0.005)
Trip duration	−0.084***	−0.085***	−0.084***	0.044***	0.044***	0.044***
	(0.011)	(0.011)	(0.011)	(0.007)	(0.007)	(0.007)
Number of flights	−0.303***	−0.306***	−0.305***	−0.035***	−0.035***	−0.035***
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Number of airlines	−0.192***	−0.195***	−0.195***	0.026**	0.024**	0.025**
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Menu size	−0.015***	−0.015***	−0.015***	−0.062***	−0.063***	−0.063***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Attraction	0.222***	0.144**	0.150***	0.759***	0.741***	0.732***
	(0.043)	(0.046)	(0.042)	(0.035)	(0.035)	(0.034)
Compromise	0.694***	−0.170	−0.298*	−0.087	−0.458***	−0.460***
	(0.141)	(0.095)	(0.117)	(0.130)	(0.093)	(0.099)
Similarity	−1.932***	−1.988***	−2.137***	−0.993***	−0.804**	−0.991***
	(0.189)	(0.353)	(0.238)	(0.140)	(0.245)	(0.146)
Attraction x Compromise	−3.082***			−2.034***		
	(0.421)			(0.515)		
Attraction x Similarity		0.123		–	0.604	
		(0.737)			(0.695)	
Compromise x Similarity			3.815			0.242
			(2.088)			(1.852)
Day of week (base = Sunday)						
Monday	0.007	0.006	0.005	0.000	0.001	0.001
	(0.016)	(0.016)	(0.016)	(0.013)	(0.013)	(0.013)
Tuesday	0.039*	0.037*	0.037*	0.019	0.020	0.020
	(0.015)	(0.015)	(0.015)	(0.013)	(0.013)	(0.013)
Wednesday	0.002	0.001	0.001	0.015	0.015	0.015
	(0.015)	(0.015)	(0.015)	(0.013)	(0.013)	(0.013)
Thursday	0.010	0.009	0.009	0.004	0.004	0.004
	(0.015)	(0.015)	(0.015)	(0.013)	(0.013)	(0.013)
Friday	0.006	0.004	0.004	0.005	0.006	0.006
	(0.015)	(0.015)	(0.015)	(0.013)	(0.013)	(0.013)
Saturday	−0.029	−0.030	−0.030	−0.009	−0.009	−0.009
	(0.017)	(0.017)	(0.017)	(0.015)	(0.015)	(0.015)
Duration of stay	0.172***	0.174***	0.174***	0.081***	0.082***	0.082***

(Continues)

TABLE G1 | (Continued)

	Full dataset			Reduced dataset		
	Model 7	**Model 8**	**Model 9**	**Model 7**	**Model 8**	**Model 9**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Departure time	−0.001	−0.001	−0.001	0.002	0.003	0.003
	(0.012)	(0.012)	(0.012)	(0.015)	(0.015)	(0.015)
Arrival time	0.032***	0.031***	0.031***	0.041***	0.041***	0.041***
	(0.008)	(0.008)	(0.008)	(0.011)	(0.011)	(0.011)
Constant	−1.432***	−1.401***	−1.398***	−0.865***	−0.856***	−0.853***
	(0.022)	(0.022)	(0.021)	(0.019)	(0.019)	(0.019)
Number of observations	344,386	344,386	344,386	40,916	40,916	40,916
Number of choices	5784	5784	5784	5017	5017	5017
Akaike information criterion	45,081	45,136	45,133	28,107	28,123	28,124
Log likelihood	−22,520	−22,548	−22,546	−14,033	−14,042	−14,042

Note: Results from random effects probit estimations. The reduced dataset is generated after discarding all alternatives not belonging to the same cluster as the chosen alternative in each menu. Clustering performed using the affinity propagation algorithm. Groups represent menus/clusters in the datasets. Robust (menu-level clustered) standard errors are in parentheses. "Menu size" describes the consideration set size in the case of reduced dataset regressions. Significance key: "***" indicates $p < 0.001$, "**" indicates $p < 0.01$, "*" indicates $p < 0.05$.

Appendix H

Logit Estimation

TABLE H1 | Regression output from logistic estimation.

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Price	−0.544***	−0.555***	−0.502***	−0.549***	−0.538***	−0.473***
	(0.015)	(0.015)	(0.020)	(0.015)	(0.015)	(0.020)
Trip duration	−0.212***	−0.256***	−0.172***	−0.255***	−0.259***	−0.160***
	(0.017)	(0.017)	(0.027)	(0.017)	(0.016)	(0.026)
Number of flights	−0.754***	−0.656***	−0.665***	−0.650***	−0.618***	−0.623***
	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Number of airlines	−0.482***	−0.480***	−0.455***	−0.478***	−0.465***	−0.434***
	(0.020)	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)
Menu size	−0.036***	−0.036***	−0.036***	−0.036***	−0.035***	−0.035***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Attraction			0.418***			0.496***
			(0.100)			(0.099)
Compromise				−0.731**		−0.401
				(0.231)		(0.231)
Similarity					−5.424***	−5.484***
					(0.492)	(0.499)
Day of week (base = Sunday)						
Monday		0.005	−0.000	0.006	−0.007	−0.013
		(0.038)	(0.038)	(0.038)	(0.037)	(0.037)
Tuesday		0.055	0.047	0.059	0.041	0.033

(Continues)

TABLE H1 | (Continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
		(0.036)	(0.036)	(0.036)	(0.035)	(0.035)
Wednesday		−0.014	−0.018	−0.012	−0.033	−0.038
		(0.036)	(0.036)	(0.036)	(0.035)	(0.035)
Thursday		−0.006	−0.007	−0.002	−0.010	−0.010
		(0.036)	(0.035)	(0.036)	(0.035)	(0.034)
Friday		−0.033	−0.035	−0.028	−0.033	−0.032
		(0.037)	(0.036)	(0.037)	(0.036)	(0.036)
Saturday		−0.055	−0.054	−0.054	−0.054	−0.053
		(0.041)	(0.041)	(0.041)	(0.040)	(0.040)
Duration of stay		0.404***	0.405***	0.402***	0.392***	0.392***
		(0.016)	(0.016)	(0.016)	(0.015)	(0.015)
Departure time		−0.033	−0.029	−0.036	−0.020	−0.018
		(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Arrival time		0.086***	0.090***	0.087***	0.074***	0.079***
		(0.019)	(0.019)	(0.019)	(0.018)	(0.018)
Constant	−2.379***	−2.452***	−2.567***	−2.415***	−2.326***	−2.442***
	(0.023)	(0.039)	(0.048)	(0.041)	(0.038)	(0.049)
Number of observations	344,386	344,386	344,386	344,386	344,386	344,386
Number of choices	5784	5784	5784	5784	5784	5784
Akaike information criterion	46,315	45,454	45,438	45,445	45,303	45,277
Log likelihood	−23,151	−22,712	−22,703	−22,706	−22,635	−22,620

Note: Results from random effects logit estimations. Groups represent menus/clusters are in the datasets. Robust (menu-level clustered) standard errors in parentheses. Significance key: "***" indicates $p < 0.001$, "**" indicates $p < 0.01$, "*" indicates $p < 0.05$.

Appendix I

Complete Regression Outputs

TABLE I1 | Complete regression output for the results reported in Table 2.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Price	−0.237***	−0.245***	−0.230***	−0.243***	−0.239***	−0.219***
	(0.007)	(0.007)	(0.009)	(0.007)	(0.008)	(0.009)
Trip duration	−0.092***	−0.111***	−0.089***	−0.111***	−0.113***	−0.084***
	(0.007)	(0.007)	(0.011)	(0.007)	(0.007)	(0.011)
Number of flights	−0.364***	−0.322***	−0.324***	−0.319***	−0.305***	−0.306***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)	(0.007)
Number of airlines	−0.209***	−0.210***	−0.203***	−0.209***	−0.205***	−0.195***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)
Menu size	−0.016***	−0.016***	−0.016***	−0.016***	−0.015***	−0.015***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Attraction			0.111**			0.147***
			(0.042)			(0.042)

(Continues)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Compromise				−0.284**		−0.171
				(0.094)		(0.094)
Similarity					−1.931***	−1.940***
					(0.187)	(0.190)
Day of week (base = Sunday)						
Monday		0.011	0.010	0.011	0.007	0.006
		(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Tuesday		0.044**	0.042**	0.046**	0.039*	0.037*
		(0.016)	(0.016)	(0.016)	(0.015)	(0.015)
Wednesday		0.009	0.008	0.010	0.002	0.001
		(0.016)	(0.016)	(0.016)	(0.015)	(0.015)
Thursday		0.010	0.010	0.012	0.009	0.009
		(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Friday		0.004	0.003	0.006	0.003	0.004
		(0.016)	(0.016)	(0.016)	(0.015)	(0.015)
Saturday		−0.031	−0.031	−0.031	−0.030	−0.030
		(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Duration of stay		0.179***	0.179***	0.178***	0.175***	0.174***
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Departure time		−0.006	−0.005	−0.007	−0.001	−0.001
		(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Arrival time		0.034***	0.036***	0.035***	0.029***	0.031***
		(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Constant	−1.385***	−1.420***	−1.451***	−1.405***	−1.372***	−1.402***
	(0.010)	(0.017)	(0.021)	(0.018)	(0.018)	(0.021)
Number of observations	344,386	344,386	344,386	344,386	344,386	344,386
Number of choices	5784	5784	5784	5784	5784	5784
Akaike information criterion	46,151	45,268	45,264	45,260	45,146	45,134
Log likelihood	−23,068	−22,618	−22,615	−22,613	−22,556	−22,548

Note: Results from random effects probit estimations. Reduced dataset is generated after discarding all alternatives not belonging to the same cluster as the chosen alternative in each menu. Clustering performed using the affinity propagation algorithm. Groups represent menus/clusters in the datasets. Robust (menu-level clustered) standard errors are in parentheses. Significance key: "***" indicates $p < 0.001$, "**" indicates $p < 0.01$, "*" indicates $p < 0.05$.

TABLE I2 | Complete regression output for the results reported in Table 3.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Price	−0.123***	−0.129***	−0.089***	−0.128***	−0.131***	−0.091***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Trip duration	−0.032***	−0.053***	−0.046***	−0.049***	−0.054***	−0.044***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Number of flights	−0.002	−0.024***	−0.034***	−0.021**	−0.024***	−0.035***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Number of airlines	−0.016*	−0.018*	−0.022**	−0.017*	−0.023**	−0.025**
	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Consideration set size	−0.061***	−0.060***	−0.064***	−0.059***	−0.059***	−0.063***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Attraction			0.761***			0.732***
			(0.033)			(0.034)
Compromise				−0.695***		−0.457***
				(0.092)		(0.093)
Similarity					−1.080***	−0.985***
					(0.141)	(0.140)
Day of week (base = Sunday)						
Monday		0.002	−0.001	0.006	0.002	0.001
		(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Tuesday		0.028*	0.021	0.035**	0.021*	0.020
		(0.012)	(0.013)	(0.012)	(0.013)	(0.013)
Wednesday		0.025**	0.017	0.031**	0.019	0.015
		(0.012)	(0.013)	(0.012)	(0.013)	(0.013)
Thursday		0.006	0.005	0.013	0.000	0.005
		(0.012)	(0.012)	(0.012)	(0.012)	(0.013)
Friday		0.003	0.006	0.008	−0.000	0.006
		(0.012)	(0.012)	(0.012)	(0.013)	(0.013)
Saturday		−0.015	−0.011	−0.010	−0.015	−0.009
		(0.014)	(0.015)	(0.014)	(0.014)	(0.015)
Duration of stay		0.080***	0.084***	0.078***	0.080***	0.082***
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Departure time		−0.021	−0.001	−0.023	−0.015	0.003
		(0.014)	(0.015)	(0.014)	(0.014)	(0.015)
Arrival time		0.040***	0.041***	0.0436***	0.037***	0.041***
		(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Constant	−0.667***	−0.726***	−0.880***	−0.716***	−0.711***	−0.853***
	(0.011)	(0.016)	(0.019)	(0.016)	(0.016)	(0.019)
Number of observations	40,916	40,916	40,916	40,916	40,916	40,916
Number of choices	5017	5017	5017	5017	5017	5017
Akaike information criterion	28,771	28,674	28,176	28,611	28,637	28,122
Log likelihood	−14,379	−14,321	−14,071	−14,289	−14,301	−14,042

Note: Results from random effects probit estimations. The reduced dataset is generated after discarding all alternatives not belonging to the same cluster as the chosen alternative in each menu. Clustering performed using the affinity propagation algorithm. Groups represent menus/clusters in the datasets. Robust (menu-level clustered) standard errors are in parentheses. Significance key: "***" indicates $p < 0.001$, "**" indicates $p < 0.01$, "*" indicates $p < 0.05$.

TABLE I3 | Complete regression output for models from Table 2 estimated on dataset after dropping all menus where the cluster to which the chosen alternative belongs does not contain at least three alternatives.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Price	−0.227***	−0.234***	−0.211***	−0.233***	−0.229***	−0.200***
	(0.007)	(0.007)	(0.009)	(0.007)	(0.007)	(0.009)
Trip duration	−0.102***	−0.126***	−0.092***	−0.126***	−0.129***	−0.087***
	(0.007)	(0.008)	(0.012)	(0.008)	(0.007)	(0.012)
Number of flights	−0.344***	−0.299***	−0.301***	−0.297***	−0.284***	−0.285***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Number of airlines	−0.208***	−0.208***	−0.197***	−0.207***	−0.203***	−0.190***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Menu size	−0.014***	−0.015***	−0.014***	−0.015***	−0.014***	−0.014***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Attraction			0.172***			0.209***
			(0.045)			(0.044)
Compromise				−0.205*		−0.0954
				(0.096)		(0.097)
Similarity					−1.739***	−1.787***
					(0.190)	(0.193)
Day of week (base = Sunday)						
Monday		0.001	−0.001	0.001	−0.001	−0.003
		(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Tuesday		0.029	0.026	0.030*	0.026	0.023
		(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Wednesday		0.001	−0.000	0.001	−0.005	−0.006
		(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Thursday		0.002	0.003	0.003	0.003	0.003
		(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Friday		−0.002	−0.002	−0.001	−0.002	−0.001
		(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Saturday		−0.024	−0.023	−0.023	−0.023	−0.022
		(0.016)	(0.016)	(0.017)	(0.016)	(0.016)
Duration of stay		0.184***	0.184***	0.183***	0.180***	0.179***
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Departure time		−0.002	−0.000	−0.003	0.004	0.005
		(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Arrival time		0.032***	0.034***	0.032***	0.027***	0.029***
		(0.008)	(0.008)	(0.008)	(0.008)	(0.009)
Constant	−1.445***	−1.471***	−1.519***	−1.459***	−1.428***	−1.479***
	(0.010)	(0.017)	(0.021)	(0.018)	(0.017)	(0.022)
Number of observations	311,364	311,364	311,364	311,364	311,364	311,364
Number of choices	5017	5017	5017	5017	5017	5017
Akaike information criterion	42,129	41,324	41,312	41,321	41,238	41,221
Log likelihood	−21,058	−20,646	−20,639	−20,643	−20,602	−20,592

Note: Results from random effects probit estimations. The reduced dataset is generated after discarding all alternatives not belonging to the same cluster as the chosen alternative in each menu. Clustering performed using the affinity propagation algorithm. Groups represent menus/clusters in the datasets. Robust (menu-level clustered) standard errors are in parentheses. Significance key: "***" indicates $p < 0.001$, "**" indicates $p < 0.01$, "*" indicates $p < 0.05$.

TABLE I4 | Complete regression output for the results reported in Table B1.

	Model 5	Model 6
Price	−0.243***	−0.225***
	(0.007)	(0.009)
Trip duration	−0.112***	−0.088***
	(0.007)	(0.011)
Number of flights	−0.315***	−0.314***
	(0.007)	(0.007)
Number of airlines	−0.208***	−0.199***
	(0.009)	(0.009)
menusize	−0.016***	−0.016***
	(0.000)	(0.000)
Attraction		0.119*
		(0.042)
Compromise		−0.288*
		(0.094)
Similarity	−3.114***	−3.157***
	(0.519)	(0.522)
Day of week (base = Sunday)		
Monday	0.006	0.005
	(0.016)	(0.016)
Tuesday	0.036	0.035
	(0.015)	(0.015)
Wednesday	0.001	0.000
	(0.015)	(0.015)
Thursday	0.005	0.007
	(0.015)	(0.015)
Friday	−0.001	0.001
	(0.015)	(0.015)
Saturday	−0.031	−0.030
	(0.017)	(0.017)
Duration of stay	0.178***	0.178***
	(0.007)	(0.007)
Departure time	−0.001	−0.001
	(0.012)	(0.012)
Arrival time	0.031***	0.032***
	(0.008)	(0.008)
Constant	−1.390***	−1.406***
	(0.017)	(0.021)
Observations	344,386	344,386
Number of individual	5784	5784

(Continues)

TABLE I4 | (Continued)

	Model 5	Model 6
Akaike information criterion	45,196	45,182
Log likelihood	−22,582	−22,572

Note: Results from random effects probit estimations. Robust (menu-level clustered) standard errors are in parentheses. Significance key: "***" indicates $p < 0.001$, "**" indicates $p < 0.01$, "*" indicates $p < 0.05$.

TABLE I5 | Complete regression output for results reported in Table C1.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Price	−0.237***	−0.245***	−0.204***	−0.240***	−0.240***	−0.194***
	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)
Trip duration	−0.092***	−0.109***	−0.047***	−0.109***	−0.108***	−0.046***
	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)
Number of flights	−0.364***	−0.320***	−0.321***	−0.313***	−0.306***	−0.302***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Number of airlines	−0.209***	−0.209***	−0.178***	−0.211***	−0.208***	−0.179***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
menusize	−0.015***	−0.016***	−0.016***	−0.016***	−0.016***	−0.017***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Attraction			0.005***			0.005***
			(0.001)			(0.001)
Compromise				−0.015***		−0.014***
				(0.002)		(0.002)
Similarity					−0.020***	−0.018***
					(0.003)	(0.003)
Day of week (base = Sunday)						
Monday		0.270	0.271	0.268	0.270	0.270
		(0.139)	(0.138)	(0.139)	(0.139)	(0.138)
Tuesday		0.269	0.274	0.267	0.269	0.272
		(0.154)	(0.153)	(0.154)	(0.154)	(0.153)
Wednesday		0.093	0.104	0.084	0.091	0.094
		(0.167)	(0.165)	(0.167)	(0.167)	(0.165)
Thursday		−0.060	−0.038	−0.058	−0.070	−0.044
		(0.175)	(0.172)	(0.174)	(0.175)	(0.172)
Friday		−0.103	−0.086	−0.100	−0.108	−0.087
		(0.161)	(0.159)	(0.160)	(0.162)	(0.159)
Saturday		−0.227	−0.220	−0.223	−0.234	−0.222
		(0.139)	(0.137)	(0.138)	(0.140)	(0.137)
Duration of stay		0.189***	0.189***	0.189***	0.184***	0.183***
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Departure time		−0.003	−0.004	−0.000	0.005	0.005
		(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Arrival time		0.033**	0.032**	0.038**	0.039**	0.041**
		(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Constant	−2.379***	−2.452***	−2.567***	−2.415***	−2.326***	−2.442***
	(0.023)	(0.039)	(0.048)	(0.041)	(0.038)	(0.049)
Number of observations	344,386	344,386	344,386	344,386	344,386	344,386
Number of choices	5784	5784	5784	5784	5784	5784
Akaike information criterion	46,315	45,454	45,438	45,445	45,303	45,277
Log likelihood	−23,151	−22,712	−22,703	−22,707	−22,635	−22,620

Note: Results from probit estimations after de-meaning data on menu the menu level. Robust (menu-level clustered) standard errors are in parentheses. Significance key: "***" indicates $p < 0.001$, "**" indicates $p < 0.01$, "*" indicates $p < 0.05$.