Choice Set Optimization Under Discrete Choice Models of Group Decisions

Kiran Tomlinson [1] **Austin R. Benson** [1]

Abstract

The way that people make choices or exhibit preferences can be strongly affected by the set of available alternatives, often called the choice set. Furthermore, there are usually heterogeneous preferences, either at an individual level within small groups or within sub-populations of large groups. Given the availability of choice data, there are now many models that capture this behavior in order to make effective predictions—however, there is little work in understanding how directly changing the choice set can be used to influence the preferences of a collection of decision-makers. Here, we use discrete choice modeling to develop an optimization framework of such interventions for several problems of group influence, namely maximizing agreement or disagreement and promoting a particular choice. We show that these problems are NP-hard in general, but imposing restrictions reveals a fundamental boundary: promoting a choice can be easier than encouraging consensus or sowing discord. We design approximation algorithms for the hard problems and show that they work well on real-world choice data.

1. Context effects and optimizing choice sets

Choosing from a set of alternatives is one of the most important actions people take, and choices determine the composition of governments, the success of corporations, and the formation of social connections. For these reasons, choice models have received significant attention in the fields of economics (Train, 2009), psychology (Tversky & Kahneman, 1981), and, as human-generated data has become increasingly available online, computer science (Overgoor et al., 2019; Seshadri et al., 2019; Rosenfeld et al., 2020). In many cases, it is important that people have heterogeneous preferences; for example, people living in different parts of a town might prefer different government policies.

Much of the computational work on choice has been devoted to fitting models for predicting future choices. In addition to prediction, another area of interest is determining effective interventions to influence choice—advertising and political campaigning are prime examples. In heterogeneous groups, the goal might be to encourage consensus (Amir et al., 2015), or, for an ill-intentioned adversary, to sow discord, e.g., amongst political parties (Rosenberg et al., 2020).

One particular method of influence is introducing new alternatives or options. While early economic models assume that alternatives are irrelevant to the relative ranking of options (Luce, 1959; McFadden, 1974), experimental work has consistently found that new alternatives have strong effects on our choices (Huber et al., 1982; Simonson & Tversky, 1992; Shafir et al., 1993; Trueblood et al., 2013). These effects are often called *context effects* or *choice set effects*. A well-known example is the compromise effect (Simonson, 1989), which describes how people often prefer a middle ground (e.g., the middle-priced wine). Direct measurements on choice data have also revealed choice set effects in several domains (Benson et al., 2016; Seshadri et al., 2019).

Here, we pose adding new alternatives as a discrete optimization problem for influencing a collection of decision makers, such as the inhabitants of a city or the visitors to a website. To this end, we consider various models for how someone makes a choice from a given set of alternatives, where the model parameters can be readily estimated from data. In our setup, everyone has a base set of alternatives from which they make a choice, and the goal is to find a set of additional alternatives to optimize some function of the group's joint preferences on the base set. We specifically analyze three objectives: (i) *agreement* in preferences amongst the group; (ii) *disagreement* in preferences amongst the group; and (iii) *promotion* of a particular item (decision).

We use the framework of *discrete choice* (Train, 2009) to probabilistically model a person's choice from a given set of items, called the *choice set*. These models are parameterized for individual preferences, and when fitting parameters from data, preferences are commonly aggregated at the level of a sub-population of individuals. Discrete choice models such as the multinomial logit and elimination-by-aspects have played a central role in behavioral economics for several

[1]Department of Computer Science, Cornell University, Ithaca, New York, USA. Correspondence to: Kiran Tomlinson <kt@cs.cornell.edu>, Austin R. Benson <arb@cs.cornell.edu>.

Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020.

decades with diverse applications, including forest management (Hanley et al., 1998), social networks formation (Overgoor et al., 2019), and marketing campaigns (Fader & McAlister, 1990). More recently, new choice data and algorithms have spurred machine learning research on models for choice set effects (Ragain & Ugander, 2016; Chierichetti et al., 2018b; Seshadri et al., 2019; Pfannschmidt et al., 2019; Rosenfeld et al., 2020; Bower & Balzano, 2020).

We provide the relevant background on discrete choice models in Section 2. From this, we formally define and analyze three choice set optimization problems—AGREEMENT, DISAGREEMENT, and PROMOTION—and analyze them under four discrete choice models: multinomial logit (McFadden, 1974), the context dependent random utility model (Seshadri et al., 2019), nested logit (McFadden, 1978), and elimination-by-aspects (Tversky, 1972). We first prove that the choice set optimization problems are NP-hard in general for these models. After, we identify natural restrictions of the problems under which they become tractable. These restrictions reveal a fundamental boundary: promoting a particular item within a group is easier than minimizing or maximizing consensus. More specifically, we show that restricting the choice models can make PROMOTION tractable while leaving AGREEMENT and DISAGREEMENT NP-hard, indicating that the interaction between individuals introduces significant complexity to choice set optimization.

After this, we provide efficient approximation algorithms with guarantees for all three problems under several choice models, and we validate our algorithms on choice data. Model parameters are learned for different types of individuals based on features (e.g., where someone lives). From these learned models, we apply our algorithms to optimize group-level preferences. Our algorithms outperform a natural baseline on real-world data coming from transportation choices, insurance policy purchases, and online shopping.

1.1. Related work

Our work fits within recent interest from computer science and machine learning on discrete choice models in general and choice set effects in particular. For example, choice set effects abundant in online data has led to richer data models (Ieong et al., 2012; Chen & Joachims, 2016; Ragain & Ugander, 2016; Seshadri et al., 2019; Makhijani & Ugander, 2019; Rosenfeld et al., 2020; Bower & Balzano, 2020), new methods for testing the presence of choice set effects (Benson et al., 2016; Seshadri et al., 2019; Seshadri & Ugander, 2019), and new learning algorithms (Kleinberg et al., 2017; Chierichetti et al., 2018b). More broadly, there are efforts on learning algorithms for multinomial logit mixtures (Oh & Shah, 2014; Ammar et al., 2014; Kallus & Udell, 2016; Zhao & Xia, 2019), Plackett-Luce models (Maystre & Grossglauser, 2015; Zhao et al., 2016), and other random utility models (Oh et al., 2015; Chierichetti et al., 2018a; Benson et al., 2018).

One of our optimization problems is maximizing group agreement by introducing new alternatives. This is motivated in part by how additional context can sway opinion on controversial topics (Munson et al., 2013; Liao & Fu, 2014; Graells-Garrido et al., 2014). There are also related algorithms for decreasing polarization in social networks (Garimella et al., 2017; Matakos et al., 2017; Chen et al., 2018; Musco et al., 2018), although we have no explicit network and adopt a choice-theoretic framework.

Our choice set optimization framework is similar to assortment optimization in operations research, where the goal is find the optimal set of products to offer in order to maximize revenue (Talluri & Van Ryzin, 2004). Discrete choice models are extensively used in this line of research, including the multinomial logit (Rusmevichientong et al., 2010; 2014) and nested logit (Gallego & Topaloglu, 2014; Davis et al., 2014) models. We instead focus our attention primarily on optimizing agreement among individuals, which has not been explored in traditional revenue-focused assortment optimization.

Finally, our problems relate to group decision-making. In psychology, introducing new shared information is critical for group decisions (Stasser & Titus, 1985; Lu et al., 2012). In computer science, the complexity of group Bayesian reasoning is a concern (Hązła et al., 2017; 2019).

2. Background and preliminaries

We first introduce the discrete choice models that we analyze. In the setting we explore, one or more individuals make a (possibly random) *choice* of a single item (or alternative) from a finite set of items called a *choice set*. We use $\mathcal{U}$ to denote the universe of items and $C \subseteq \mathcal{U}$ the choice set. Thus, given C, an individual chooses some item $x \in C$.

Given C, a discrete choice *model* provides a probability for choosing each item $x \in C$. We analyze four broad discrete choice models that are all *random utility models* (RUMs), which derive from economic rationality. In a RUM, an individual observes a random utility for each item $x \in C$ and then chooses the one with the largest utility. We model each individual's choices through the same RUM but with possibly different parameters to capture preference heterogeneity. In this sense, we have a mixture model.

Choice data typically contains many observations from various choice sets. We occasionally have data specific enough to model the choices of a particular individual, but often only one choice is recorded per person, making accurate preference learning impossible at that scale. Thus, we instead model the heterogeneous preferences of sub-populations or categories of individuals. For convenience, we still use "individual" or "person" when referring to components of a mixed population, since we can treat each component as a decision-making agent with its own preferences. In contrast, we use the term "group" to refer to the entire population. We use A to denote the set of individuals (in the broad sense above), and $a \in A$ indexes model parameters.

The parameters of the RUMs we analyze can be inferred from data, and our theoretical results and algorithms assume that we have learned these parameters. Our analysis focuses on how the probability of selecting an item x from a choice set C changes as we add new alternative items from $\overline{C} = \mathcal{U} \setminus C$ to the choice set.

We let $n = |A|$, $k = |C|$, and $m = |\overline{C}|$ for notation. We mostly use $n = 2$, which is sufficient for hardness proofs.

Multinomial logit (MNL). The multinomial logit (MNL) model (McFadden, 1974) is the workhorse of discrete choice theory. In MNL, an individual a's preferences are encoded by a true utility $u_a(x)$ for every item $x \in \mathcal{U}$. The observations are noisy random utilities $\tilde{u}_a(x) = u_a(x) + \varepsilon$, where ε follows a Gumbel distribution. Under this model, the probability that individual a picks item x from choice set C (i.e., $x = \arg\max_{y \in C} \tilde{u}_a(y)$) is the softmax over item utilities:

$$\Pr(a \leftarrow x \mid C) = \frac{e^{u_a(x)}}{\sum_{y \in C} e^{u_a(y)}}. \tag{1}$$

We use the term *exp-utility* for terms like $e^{u_a(x)}$. The utility of an item is often parameterized as a function of features of the item in order to generalize to unseen data. For example, a linear function is an additive utility model (Tversky & Simonson, 1993) and looks like logistic regression. In our analysis, we work directly with the utilities.

The MNL satisfies *independence of irrelevant alternatives* (IIA) (Luce, 1959), the property that for any two choice sets C, D and two items $x, y \in C \cap D$: $\frac{\Pr(a \leftarrow x \mid C)}{\Pr(a \leftarrow y \mid C)} = \frac{\Pr(a \leftarrow x \mid D)}{\Pr(a \leftarrow y \mid D)}$. In other words, the choice set has no effect on a's relative probability of choosing x or y.[1] Although IIA is intuitively pleasing, behavioral experiments show that it is often violated in practice (Huber et al., 1982; Simonson & Tversky, 1992). Thus, there are many models that account for IIA violations, including the other ones we analyze.

Context-dependent random utility model (CDM). The CDM (Seshadri et al., 2019) is an extension of MNL that can model IIA violations. The core idea is to approximate choice set effects by the effect of each item's presence on the utilities of the other items. For instance, a diner's preference for a ribeye steak may decrease relative to a fish option if filet mignon is also available. Formally, each item z exerts a pull on a's utility from x, which we denote $p_a(z, x)$. The CDM then resembles the MNL with utilities $u_a(x \mid C) = u_a(x) + \sum_{z \in C} p_a(z, x)$. This leads to choice probabilities that are a softmax over the context-dependent utilities:

$$\Pr(a \leftarrow x \mid C) = \frac{e^{u_a(x \mid C)}}{\sum_{y \in C} e^{u_a(y \mid C)}}. \tag{2}$$

Nested logit (NL). The nested logit (NL) model (McFadden, 1978) instead accounts for choice set effects by grouping similar items into *nests* that people choose between successively. For example, a diner may first choose between a vegetarian, fish, or steak meal and then select a particular dish. NL can be derived by introducing correlation between the random utility noise ε in MNL; here, we instead consider a generalized tree-based version of the model.[2]

The (generalized) NL model for an individual a consists of a tree T_a with a leaf for each item in $\mathcal{U}$, where the internal nodes represent categories of items. Rather than having a utility only on items, each person a also has utilities $u_a(v)$ on all nodes $v \in T_a$ (except the root). Given a choice set C, let $T_a(C)$ be the subtree of T_a induced by C and all ancestors of C. To choose an item from C, a starts at the root and repeatedly picks between the children of the current node according to the MNL model until reaching a leaf.

Elimination-by-aspects (EBA). While the previous models are based on MNL, the elimination-by-aspects (EBA) model (Tversky, 1972) has a different structure. In EBA, each item x has a set of *aspects* x' representing properties of the item, and person a has a utility $u_a(\chi) > 0$ on each aspect χ. An item is chosen by repeatedly picking an aspect with probability proportional to its utility and eliminating all items that do not have that aspect until only one item remains (or, if all remaining items have the same aspects, the choice is made uniformly at random). For example, a diner may first eliminate items that are too expensive, then disregard meat options, and finally look for dishes with pasta before choosing mushroom ravioli. Formally, let $C' = \bigcup_{x \in C} x'$ be the set of aspects of items in C and let $C^0 = \bigcap_{x \in C} x'$ be the aspects shared by all items in C. Additionally, let $C_\chi = \{x \in C \mid \chi \in x'\}$. The probability that individual a picks item x from choice set C is recursively defined as

$$\Pr(a \leftarrow x \mid C) = \frac{\sum_{\chi \in x' \setminus C^0} u_a(\chi) \Pr(a \leftarrow x \mid C_\chi)}{\sum_{\psi \in C' \setminus C^0} u_a(\psi)}. \tag{3}$$

If all remaining items have the same aspects ($C' = C^0$), the denominator is zero, and $\Pr(a \leftarrow x \mid C) = \frac{1}{|C|}$ in that case.

Encoding MNLs in other models. Although the three models with context effects appear quite different, they all subsume the MNL model. Thus, if we prove a problem hard under MNL, then it is hard under all four models.

[1] Over $a \in A$, we have a mixed logit which does not have to satisfy IIA (McFadden & Train, 2000). Here, we are interested in the IIA property at the individual level.

[2] Certain parameter regimes in this generalized model do not correspond to RUMs (Train, 2009), but this model is easier to analyze and captures the salient structure.

Lemma 1. *The MNL model is a special case of the CDM, NL, and EBA models.*

Proof. Let $\mathcal{M}$ be an MNL model. For the CDM, use the utilities from $\mathcal{M}$ and set all pulls to 0. For NL, make all items children of T_a's root and use the utilities from $\mathcal{M}$. Lastly, for EBA, assign a unique aspect χ_x to each item $x \in \mathcal{U}$ with utility $u_a(\chi_x) = e^{u_a(x)}$. Following (3), $\Pr(a \leftarrow x \mid C) = \frac{u_a(\chi_x)\Pr(a \leftarrow x \mid C_{\chi_x})}{\sum_{\psi \in C' \setminus C^0} u_a(\psi)}$. Since $C_{\chi_x} = \{x\}$, $\Pr(a \leftarrow x \mid C_{\chi_x}) = 1$ and thus $\Pr(a \leftarrow x \mid C) \propto u_a(\chi_x) = e^{u_a(x)}$, matching the MNL $\mathcal{M}$. $\square$

3. Choice set optimization problems

By introducing new alternatives to the choice set C, we can modify the relationships amongst individual preferences, resulting in different dynamics at the collective level. Similar ideas are well-studied in voting models, e.g., introducing alternatives to change winners selected by Borda count (Easley & Kleinberg, 2010). Here, we study how to optimize choice sets for various group-level objectives, measured in terms of individual choice probabilities coming from discrete choice models.

Agreement and Disagreement. Since we are modeling the preferences of a collection of decision-makers, one important metric is the amount of disagreement (conversely, agreement) about which item to select. Given a set of alternatives $Z \subseteq \overline{C}$ we might introduce, we quantify the disagreement this would induce as the sum of all pairwise differences between individual choice probabilities over C:

$$D(Z) = \sum_{\{a,b\} \subseteq A, x \in C} |\Pr(a \leftarrow x \mid C \cup Z) - \Pr(b \leftarrow x \mid C \cup Z)|. \tag{4}$$

Here, we care about the disagreement on the original choice set C that results from preferences over the new choice set $C \cup Z$. In this setup, C could represent core options (e.g., two major health care policies under deliberation) and Z additional alternatives designed to sway opinions.

Concretely, we study the following problem: given $A, C, \overline{C}$, and a choice model, minimize (or maximize) $D(Z)$ over $Z \subseteq \overline{C}$. We call the minimization problem AGREEMENT and the maximization problem DISAGREEMENT. AGREEMENT has applications in encouraging consensus, while DISAGREEMENT yields insight into how susceptible a group may be to an adversary who wishes to increase conflict. Another potential application for DISAGREEMENT is to enrich the diversity of preferences present in a group.

Promotion. Promoting an item is another natural objective, which is of considerable interest in online advertising and content recommendation. Given $A, C, \overline{C}$, a choice model, and a target item $x^* \in C$, the PROMOTION problem is to find the set of alternatives $Z \subseteq \overline{C}$ whose introduction maximizes the number of individuals whose "favorite" item in C is x^*. Formally, this means maximizing the number of individuals $a \in A$ for whom $\Pr(a \leftarrow x^* \mid C \cup Z) > \Pr(a \leftarrow x \mid C \cup Z)$, $x \in C$, $x \neq x^*$. This also has applications in voting, where questions about the influence of new candidates constantly arise.

One of our contributions in this paper is showing that promotion can be easier (in a computational complexity sense) than agreement or disagreement optimization.

4. Hardness results

We now characterize the computational complexity of AGREEMENT, DISAGREEMENT, and PROMOTION under the four discrete choice models. We first show that AGREEMENT and DISAGREEMENT are NP-hard under all four models and that PROMOTION is NP-hard under the three models with context effects. After, we prove that imposing additional restrictions on these discrete choice models can make PROMOTION tractable while leaving AGREEMENT and DISAGREEMENT NP-hard. The parameters of some choice models have extra degrees of freedom, e.g., MNL has additive-shift-invariant utilities. For inference, we use a standard form (e.g., sum of utilities equals zero). For ease of analysis, we do not use such standard forms, but the choice probabilities remain unambiguous.

4.1. AGREEMENT

Although the MNL model does not have any context effects, introducing alternatives to the choice set can still affect the relative preferences of two different individuals. In particular, introducing alternatives can impact disagreement in a sufficiently complex way to make identifying the optimal set of alternatives computationally hard. Our proof of Theorem 1 uses a very simple MNL in the reduction, with only two individuals and two items in C, where the two individuals have *exactly the same utilities on alternatives*. In other words, even when individuals agree about new alternatives, encouraging them to agree over the choice set is hard.

Theorem 1. *In the MNL model,* AGREEMENT *is NP-hard, even with just two items in C and two individuals that have identical utilities on items in $\overline{C}$.*

Proof. By reduction from PARTITION, an NP-complete problem (Karp, 1972). Let S be the set of integers we wish to partition into two subsets with equal sum. We construct an instance of DISAGREEMENT with $A = \{a, b\}$, $C = \{x, y\}$, $\overline{C} = S$ (abusing notation to identify alternatives with the PARTITION integers). Let $t = \frac{1}{2}\sum_{z \in S} z$. Define the utilities as: $u_a(x) = \log t$, $u_b(x) = \log 3t$, $u_a(y) = \log t$, $u_b(y) = \log 2t$, and $u_a(z) = u_b(z) = \log z$ for all $z \in \overline{C}$. The disagreement induced by a set of alternatives $Z \subseteq \overline{C}$ is characterized by its sum of exp-utility $s_Z = \sum_{z \in Z} z$:

$$D(Z) = \frac{t}{2t+s_Z} - \frac{3t}{5t+s_Z} + \frac{t}{2t+s_Z} - \frac{2t}{5t+s_Z}.$$

The total exp-utility of all items in $\overline{C}$ is $2t$. On the interval $[0, 2t]$, $D(Z)$ is minimized at $s_Z = t$ (Appendix A.1; Fig. 4, left). Thus, if we could efficiently find the set Z minimizing $D(Z)$, then we could efficiently solve PARTITION. □

From Lemma 1, the other models we consider can all encode any MNL instance, which leads to the following corollary.

Corollary 1. AGREEMENT *is NP-hard in the CDM, NL, and EBA models.*

4.2. DISAGREEMENT

Using a similar strategy, we can construct an MNL instance whose disagreement is maximized rather than minimized at a particular target value (Theorem 2). The reduction requires an even simpler MNL setup.

Theorem 2. *In the MNL model,* DISAGREEMENT *is NP-hard, even with just one item in C and two individuals that have identical utilities on items in $\overline{C}$.*

Proof. By reduction from SUBSET SUM (Karp, 1972). Let S be a set of positive integers with target t. Let $A = \{a, b\}$, $C = \{x\}$, $\overline{C} = S$, with utilities: $u_a(x) = \log 2t$, $u_b(x) = \log t/2$, and $u_a(z) = u_b(z) = \log z$ for all $z \in \overline{C}$. Letting $s_Z = \sum_{z \in Z} z$, including $Z \subseteq \overline{C}$ makes the disagreement

$$D(Z) = \frac{2t}{2t+s_Z} - \frac{t/2}{t/2+s_Z}.$$

For $s_Z \geq 0$, $D(Z)$ is maximized at $s_Z = t$ (Appendix A.1; Fig. 4, right). Thus, if we could efficiently maximize $D(Z)$, then we could efficiently solve SUBSET SUM. □

By Lemma 1, we again have the following corollary.

Corollary 2. DISAGREEMENT *is NP-hard in the CDM, NL, and EBA models.*

4.3. PROMOTION

In choice models with no context effects, PROMOTION has a constant-time solution: under IIA, the presence of alternatives has no effect on an individual's relative preference for items in C. However, PROMOTION is more interesting with context effects, and we show that it is NP-hard for CDM, NL, and EBA. In Section 4.4, we will show that restrictions of these models make PROMOTION tractable but keep AGREEMENT and DISAGREEMENT hard.

Theorem 3. *In the CDM model,* PROMOTION *is NP-hard, even with just one individual and three items in C.*

Proof. By reduction from SUBSET SUM. Let set S with target t be an instance of SUBSET SUM. Let $A = \{a\}$, $C = \{x^*, w, y\}$, $\overline{C} = S$. Using tuples interpreted entry-wise for brevity, suppose that we have the following utilities: $u_a(\langle x^*, w, y\rangle \mid C) = \langle 1, t, -t\rangle$, $u_a(z) = -\infty$ for all $z \in \overline{C}$, and $p_a(z, \langle x^*, w, y\rangle) = \langle z, 0, 2z\rangle$ for all $z \in \overline{C}$. We wish to promote x^*. Let $s_Z = \sum_{z \in Z} z$. When we include the alternatives in Z, x^* is the item in C most likely to be chosen if and only if $1 + s_Z > t$ and $1 + s_Z > -t + 2s_Z$. Since s_Z and t are integers, this is only possible if $s_Z = t$. Thus, if we could efficiently promote x^*, then we could efficiently solve SUBSET SUM. □

We use the same Goldilocks strategy in our proofs for the NL and EBA models (details in Appendix A): by carefully defining utilities, we create choice instances where the optimal promotion solution is to pick just the right quantity of alternatives to increase preference for one item without overshooting. However, the NL model has a novel challenge compared to the CDM. With CDM, alternatives can increase the choice probability of an item in C, but in the NL, new alternatives only lower choice probabilities.

Theorem 4. *In the NL model,* PROMOTION *is NP-hard, even with just two individuals and two items in C.*

This construction relies on the two individuals having different tree structures. We will see in Section 4.4 that this is a necessary condition for the hardness of PROMOTION. Finally, we have the following hardness result for EBA.

Theorem 5. *In the EBA model,* PROMOTION *is NP-hard, even with just two individuals and two items in C.*

4.4. Restricted models that make promotion easier

We now show that, in some sense, PROMOTION is a fundamentally easier problem than AGREEMENT or DISAGREEMENT. Specifically, there are simple restrictions on CDM, NL, and EBA that make PROMOTION tractable but leave AGREEMENT and DISAGREEMENT NP-hard. Importantly, these restrictions still allow for choice set effects. In Appendix B, we also prove a strong restriction on the MNL model where AGREEMENT and DISAGREEMENT are tractable, but we could not find meaningful restrictions for similar results on the other models.

2-item CDM with equal context effects. The proof of Theorem 3 shows that PROMOTION is hard with only a single individual and three items in C. However, if C only has two items *and* context effects are the same (i.e., $p_a(z, \cdot)$ is the same for all $z \in \overline{C}$), then PROMOTION is tractable. The optimal solution is to include all alternatives that increase utility for x^* more than the other item, as doing so makes strict progress on promoting x^*. If individuals have different context effects or if there are more than two items, then there can be conflicts between which items should be included (see Appendix A.2 for a proof that 2-item CDM with *unequal* context effects makes PROMOTION NP-hard). Although this restriction makes PROMOTION tractable, it leaves AGREEMENT and DISAGREEMENT NP-hard: the

Algorithm 1 ε-additive approximation for AGREEMENT in the MNL model.

Input: n individuals A, k items C, m alternatives $\overline{C}$, utilities $u_a(\cdot) > 0$ for each $a \in A$. For brevity:
$e_{ax} \leftarrow e^{u_a(x)},\ s_a \leftarrow \sum_{z \in \overline{C}} e_{az},\ \delta \leftarrow \varepsilon/(2km\binom{n}{2})$
$L_0 \leftarrow$ empty n-dimensional array whose ath dimension has size $1+\lfloor \log_{1+\delta} s_a \rfloor$ (each cell can store a set $Z \subseteq \overline{C}$ and its n exp-utility sums for each individual)
Initialize $L_0[0, \dots, 0] \leftarrow (\emptyset, 0, \dots, 0)$ (n zeros)
for $i = 1$ **to** m **do**
  $z \leftarrow \overline{C}[i-1],\quad L_i \leftarrow L_{i-1}$
  for each cell of L_{i-1} containing $(Z, t_1, \dots, t_n)$ **do**
    $h \leftarrow n$-tuple w/ entries $\lfloor \log_{1+\delta}(t_j + e_{a_j z}) \rfloor, \forall j$
    if $L_i[h]$ is empty **then**
      $L_i[h] \leftarrow (Z \cup \{z\}, t_1 + e_{a_1 z}, \dots, t_n + e_{a_n z})$
$Z_m \leftarrow$ collection of all sets Z in cells of L_m
return $\arg\min_{Z \in Z_m} D(Z)$ (see Eq. (4))

proofs of Theorems 1 and 2 can be interpreted as 2-item and 1-item CDMs with equal (zero) context effects.

Same-tree NL. If we require that all individuals share the same NL tree structure, but still allow different utilities, then promotion becomes tractable. For each $z \in \overline{C}$, we can determine whether it reduces the relative choice probability of x^* based on its position in the tree: adding z decreases the relative choice probability of x^* if and only if z is a sibling of any ancestor of x^* (including x^*) or if it causes such a sibling to be added to $T_a(C)$. Thus, the solution to PROMOTION is to include all z not in those positions, which is a polynomial-time check. This restriction leaves AGREEMENT and DISAGREEMENT NP-hard via Theorems 1 and 2 as we can still encode any MNL model in a same-tree NL using the tree in which all items are children of the root.

Disjoint-aspect EBA. The following condition on aspects makes promoting x^* tractable: for all $z \in \overline{C}$, either $z' \cap x^{*\prime} = \emptyset$ or $z' \cap y' = \emptyset$ for all $y \in C, y \neq x^*$. That is, alternatives either share no aspects with x^* or share no aspects with other items in C. This prevents alternatives from cannibalizing from both x^* and its competitors. To promote x^*, we include all alternatives that share aspects with competitors of x^* but not x^* itself, which strictly promotes x^*. This condition is slightly weaker than requiring all items to have disjoint aspects, which reduces to MNL. However, this condition is again not sufficient to make AGREEMENT and DISAGREEMENT tractable, since any MNL model can be encoded in a disjoint-aspect EBA instance.

5. Approximation algorithms

Thus far, we have seen that several interesting group decision-making problems are NP-hard across standard discrete choice models. Here, we provide a positive result: we can compute arbitrarily good approximate solutions to many instances of these problems in polynomial time. We focus our analysis on Algorithm 1, which is an ε-additive approximation algorithm to AGREEMENT under MNL, with runtime polynomial in k, m, and $\frac{1}{\varepsilon}$, but exponential in n (recall that $k = |C|$, $m = |\overline{C}|$, and $n = |A|$). In contrast, brute force (testing every set of alternatives) is exponential in m and polynomial in k and n. AGREEMENT is NP-hard even with $n = 2$ (Theorem 1), so our algorithm provides a substantial efficiency improvement. We discuss how to extend this algorithm to other objectives and other choice models later in the section. Finally, we present a faster but less flexible mixed-integer programming approach for MNL AGREEMENT and DISAGREEMENT that performs very well in practice.

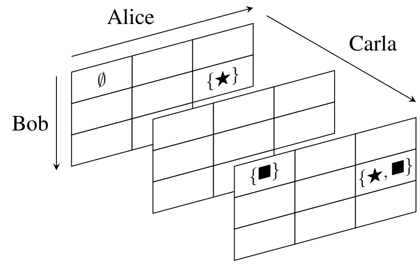

Figure 1. Example of the structure L_i used in Algorithm 1 for $n = 3$ individuals and $\overline{C} = \{\bigstar, \blacksquare\}$. Here, Alice has high utility for $\bigstar$ and low utility for $\blacksquare$, Bob has medium utility for $\bigstar$ and low utility for $\blacksquare$, and Carla has low utility for $\bigstar$ and high utility for $\blacksquare$. The exp-utility sums stored in cells are omitted.

Algorithm 1 is based on an FPTAS for SUBSET SUM (Cormen et al., 2001, Sec. 35.5), and the first parts of our analysis follow some of the same steps. The core idea of our algorithm is that a set of items can be characterized by its exp-utility sums for each individual and that there are only polynomially many combinations of exp-utility sums that differ by more than a multiplicative factor $1 + \delta$. We can therefore compute all sets of alternatives with meaningfully different impacts and pick the best one. For the purpose of the algorithm, we assume all utilities are positive (otherwise we may access a negative index); utilities can always be shifted by a constant to satisfy this requirement.

We now provide an intuitive description of Algorithm 1. The array L_i has one dimension for each individual in A (we use a hash table in practice since L_i is typically sparse). The cells along a particular dimension discretize the exp-utility sums that the individual corresponding to that dimension could have for a particular set of alternatives (Figure 1). In particular, if individual j has total exp-utility $t_j = \sum_{y \in Z} e^{u_j(y)}$ for a set Z, then we store Z at index $\lfloor \log_{1+\delta} t_j \rfloor$ along dimension j.

As the algorithm progresses, we place possible sets of alternatives Z in the cells of L_i according to their exp-utility sums $t_1, \dots, t_n$ for each individual (we store $t_1, \dots, t_n$ in the cell along with Z). We add one element at a time from $\overline{C}$

to the sets already in L_i (L_0 starts with only the empty set). If two sets have very similar exp-utility sums, they may map to the same cell, in which case only one of them is stored. If the discretization of the array is coarse enough (that is, with large enough δ), many sets of alternatives will map to the same cells, reducing the number of sets we consider and saving computational work. On the other hand, if the discretization is fine enough (δ is sufficiently small), then the best set we are left with at the end of the algorithm cannot induce a disagreement value too different from the optimal set. The proof of Theorem 6 formalizes this reasoning.

Theorem 6. *Algorithm 1 is an ε-additive approximation for* AGREEMENT *in the MNL model.*

Proof. We will use the following lemma, which says that sets mapping to the same cell have similar exp-utility sums.

Lemma 2. *Let $\overline{C}_i$ be the first i elements processed by the outer for loop. At the end of the algorithm, for all $Z \subseteq \overline{C}_i$ with exp-utility sums t_a, there exists some $Z' \in L_i$ with exp-utility sums t'_a such that $\frac{t_a}{(1+\delta)^i} < t'_a < t_a(1+\delta)^i$, for all $a \in A$ (with δ as defined in Algorithm 1, Line 2).*

The proof is in Appendix C. Now let $\beta = \varepsilon/(k\binom{n}{2})$. Following our choice of δ and using Lemma 2, at the end of the algorithm, the optimal set $Z^* \subseteq \overline{C}$ (with exp-utility sums t^*_a) has some representative Z' in L_m such that

$$\frac{t^*_a}{(1+\beta/(2m))^m} < t'_a < t^*_a\left(1+\beta/(2m)\right)^m, \ \forall a \in A.$$

Since $e^x \geq (1+x/m)^m$, we have $t^*_a/e^{\frac{\beta}{2}} < t'_a < t^*_a e^{\frac{\beta}{2}}$, and since $e^x \leq 1+x+x^2$ when $x<1$,

$$\frac{t^*_a}{1+\beta/2+\beta^2/4} < t'_a < t^*_a(1+\beta/2+\beta^2/4).$$

Finally, $\frac{t^*_a}{1+\beta} < t'_a < t^*_a(1+\beta)$ because $0<\beta<1$.

Now we show that $D(Z^*)$ and $D(Z')$ differ by at most ε. To do so, we first bound the difference between $\Pr(a \leftarrow x \mid C \cup Z^*)$ and $\Pr(a \leftarrow x \mid C \cup Z')$ by β. Let $c_a = \sum_{x \in C} e_{ax}$ be the total exp-utility of a on C. By the above reasoning,

$$\frac{e_{ax}}{c_a + t^*_a(1+\beta)} < \frac{e_{ax}}{c_a + t'_a} < \frac{e_{ax}}{c_a + \frac{t^*_a}{1+\beta}},$$

where the middle term is equal to $\Pr(a \leftarrow x \mid C \cup Z')$. From the lower bound, the difference between $\Pr(a \leftarrow x \mid C \cup Z^*)$ and $\Pr(a \leftarrow x \mid C \cup Z')$ could be as large as

$$\begin{aligned}&\frac{e_{ax}}{c_a + t^*_a} - \frac{e_{ax}}{c_a + t^*_a(1+\beta)}\\ &= \frac{e_{ax}t^*_a\beta}{(c_a+t^*_a)(c_a+t^*_a(1+\beta))} < \frac{e_{ax}t^*_a\beta}{2c_a t^*_a} \leq \frac{\beta}{2}.\end{aligned}$$

From the upper bound, the difference between $\Pr(a \leftarrow x \mid C \cup Z^*)$ and $\Pr(a \leftarrow x \mid C \cup Z')$ could be as large as

$$\begin{aligned}\frac{e_{ax}}{c_a + \frac{t^*_a}{1+\beta}} - \frac{e_{ax}}{c_a + t^*_a} &= \frac{e_{ax}t_a(1-\frac{1}{1+\beta})}{(c_a + \frac{t^*_a}{1+\beta})(c_a + t^*_a)}\\ &= \frac{e_{ax}t^*_a\beta}{(c_a(1+\beta)+t^*_a)(c_a+t^*_a)} < \frac{e_{ax}t^*_a\beta}{2c_a t^*_a} \leq \frac{\beta}{2}.\end{aligned}$$

Thus, $\Pr(a \leftarrow x \mid C \cup Z^*)$ and $\Pr(a \leftarrow x \mid C \cup Z')$ differ by at most $\frac{\beta}{2}$. Using the same argument for an individual b, the disagreement between a and b about x can only increase by β with the set Z compared to the optimal set Z^*. Since there are $\binom{n}{2}$ pairs of individuals and k items in C, the total error of the algorithm is bounded by $k\binom{n}{2}\beta = \varepsilon$. □

We now show that the runtime of Algorithm 1 is $O((m + kn^2)(1 + \lfloor\log_{1+\delta} s\rfloor)^n)$, where $s = \max_a s_a$ is the maximum exp-utility sum for any individual. Thus, for any fixed n, this runtime is bounded by a polynomial in k, m, and $\frac{1}{\varepsilon}$. To see this, first note that the size of L_i is bounded above by $(1 + \lfloor\log_{1+\delta} s\rfloor)^n$. For each $z \in \overline{C}$, we perform constant-time operations on each entry of L_i, for a total of $O(m(1 + \lfloor\log_{1+\delta} s\rfloor)^n)$ time. Then we compute $D(Z)$ for each cell of L_m, which takes $O(kn^2)$ time per cell. The total runtime is therefore $O((m + kn^2)(1 + \lfloor\log_{1+\delta} s\rfloor)^n)$, as claimed. Finally, $(1 + \lfloor\log_{1+\delta} s\rfloor)^n$ is bounded by a polynomial in m, k, and $\frac{1}{\varepsilon}$ for any fixed n (Appendix C.2).

AGREEMENT is NP-hard even when individuals have equal utilities on alternatives. In this case, we only need to compute exp-utility sums for a single individual, which brings the runtime down to $O((m + kn^2)\log_{1+\delta} s)$.

Extensions to other objectives and models. Algorithm 1 can be easily extended to any objective function that is efficiently computable from utilities. For instance, Algorithm 1 can be adapted for DISAGREEMENT by replacing the $\arg\min$ with an $\arg\max$ on Line 12.

Algorithm 1 can also be adapted for CDM and NL. The analysis is similar and details are in Appendix C, although the running times and guarantees are different. With CDM, the exponent in the runtime increases to nk for AGREEMENT and DISAGREEMENT, and the ε-additive approximation is guaranteed only if items in $\overline{C}$ exert zero pulls on each other. However, even for the general CDM, our experiments will show that the adapted algorithm remains a useful heuristic. When we adapt Algorithm 1 for NL, we retain the full approximation guarantee but the exponent in the runtime increases and has a dependence on the tree size.

PROMOTION is not interesting under MNL and also has a discrete rather than continuous objective, i.e., the number of people with favorite item x^* in C. For models with context effects, we can define a meaningful notion of approximation. We say that an item $y \in C \cup Z$ is an *ε-favorite* item of individual a if $\Pr(a \leftarrow y \mid C \cup Z) + \varepsilon \geq \Pr(a \leftarrow x \mid C \cup Z)$ for all $x \in C$. A solution then ε-approximates PROMOTION if

Table 1. Dataset statistics: item, observation, and unique choice set counts; and percent of observations in sub-population splits.

Dataset	# items	# obs.	# sets	split %
SFWork	6	5029	12	16/84
Allstate	24	97009	2697	45/55
Yoochoose	41	90493	1567	47/53

Table 2. Sum of error over all 2-item choice sets C compared to optimal (brute force) on SFWork. Algorithm 1 is optimal.

Model	Problem	Greedy	Algorithm 1
MNL	Agreement	0.03	0.00
	Disagreement	0.00	0.00
rank-2 CDM	Agreement	0.14	0.00
	Disagreement	0.13	0.00
NL	Agreement	0.00	0.00
	Disagreement	0.00	0.00

the number of people for whom x^* is an ε-favorite item is at least the value of the optimal Promotion solution. Using this, we can adapt Algorithm 1 for Promotion under CDM and NL. Again, for CDM, the approximation has guarantees in certain parameter regimes and the NL has full approximation guarantees. Since we do not have compute $D(Z)$, the runtimes loses the kn^2 term compared to the Agreement and Disagreement versions (Appendix C.5).

Finally, EBA has considerably different structure than the other models. We leave algorithms for EBA to future work.

Fast exact methods for MNL. We provide another approach for solving Agreement and Disagreement in the MNL model, based on transforming the objective functions into mixed-integer bilinear programs (MIBLPs; details in Appendix D). MIBLPs can be solved for moderate problem sizes with high-performance branch-and-bound solvers (we use Gurobi's implementation). In practice, this approach is faster than Algorithm 1 and can optimize over larger sets $\overline{C}$. However, this approach does not easily extend to CDM, NL, or Promotion and does not have a polynomial-time runtime guarantee.

6. Numerical experiments

We apply our methods to three datasets (Table 1). The SFWork dataset (Koppelman & Bhat, 2006) comes from a survey of San Francisco residents on available (choice set) and selected (choice) transportation options to get to work. We split the respondents into two segments ($|A| = 2$) according to whether or not they live in the "core residential district of San Fransisco or Berkeley." The Allstate dataset (Kaggle, 2014) consists of insurance policies (items) characterized by anonymous categorical features A–G with 2 to 4 values each. Each customer views a set of policies (the choice set) before purchasing one. We reduce the number

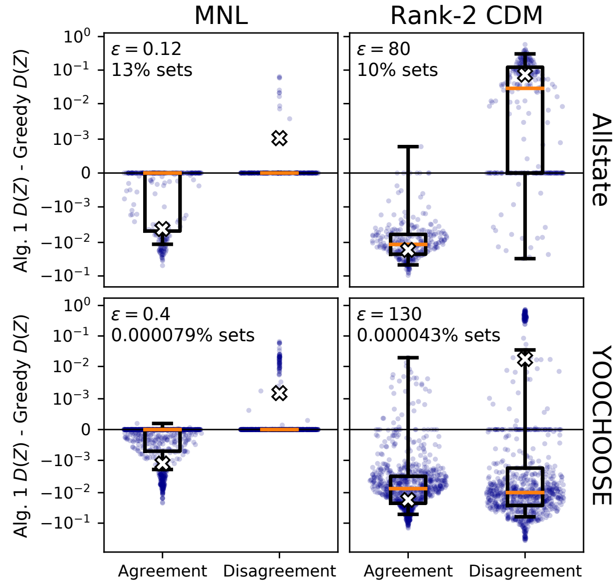

Figure 2. Algorithm 1 vs. Greedy performance box plots when applied to all 2-item choice sets in Allstate and Yoochoose under MNL and CDM (subplots also show ε and the percent of subsets of $\overline{C}$ computed by Algorithm 1, written X% sets). Each point is the difference in $D(Z)$ when Algorithm 1 and Greedy are run on a particular choice set. Horizontal spread shows approximate density and the Xs mark means. A negative (resp. positive) y-value for Agreement (resp. Disagreement) indicates that Algorithm 1 outperformed Greedy. Algorithm 1 performs better in all cases except for Disagreement under CDM on Yoochoose. Even in this exception, though, our approach finds a few very good solutions and Algorithm 1 has better mean performance.

of items to 24 by considering only features A, B, and C. To model different types of individuals, we split the data into homeowners and non-homeowners (again, $|A| = 2$). The Yoochoose dataset (Ben-Shimon et al., 2015) contains online shopping data of clicks and purchases of categorized items in user browsing sessions. Choice sets are unique categories browsed in a session and the choice is the category of the purchased product (categories appearing fewer than 20 times were omitted). We split the choice data into two subpopulations by thresholding on the purchase timestamps.

For inferring maximum-likelihood models from data, we use PyTorch's Adam optimizer (Kingma & Ba, 2015; Paszke et al., 2019) with learning rate 0.05, weight decay 0.00025, batch size 128, and the `amsgrad` flag (Reddi et al., 2018). We use the low-rank (rank-2) CDM (Seshadri et al., 2019) that expresses pulls as the inner product of item embeddings. Our code and data are available at `https://github.com/tomlinsonk/choice-set-opt`.

For SFWork under the MNL, CDM, and NL models, we considered all 2-item choice sets C (using all other items for $\overline{C}$) for Agreement and Disagreement (for the NL model, we used the best-performing tree from Koppelman & Bhat (2006)). We compare Algorithm 1 ($\varepsilon = 0.01$) to a greedy approach (henceforth, "Greedy") that builds Z by repeatedly selecting the item from $\overline{C}$ that, when added to Z, most improves the objective, if such an item exists. This dataset was small enough to compare against the optimal, brute-force solution (Table 2). In all cases, Algorithm 1 finds the optimal solution, while Greedy is often suboptimal. However, for this value of ε, we find that Algorithm 1 searches the entire space and actually computes the brute force solution (we get the number of sets analyzed by Algorithm 1 from $|L_m|$ for a given ε and compare it to $2^{|\overline{C}|}$). Even though we have an asymptotic polynomial runtime guarantee, for small enough datasets, we might not see computational savings. Running with larger ε yielded similar results, even for $\varepsilon > 2$, when our bounds are vacuous.

The results still highlight two important points. First, even on small datasets, Greedy can be sub-optimal. For example, for AGREEMENT under CDM with $C = \{\text{drive alone}, \text{transit}\}$, Algorithm 1 found the optimal $Z = \{\text{bike}, \text{walk}\}$, inferring that both sub-populations agree on both driving less and taking transit less. However, Greedy just introduced a carpool option, which has a lower effect on discouraging driving alone or taking transit, resulting in lower agreement between city and suburban residents.

Second, our theoretical bounds can be more pessimistic than what happens in practice. Thus, we can consider larger values of ε to reduce the search space; Algorithm 1 remains a principled heuristic, and we can measure how much of the search space Algorithm 1 considers. This is the approach we take for the ALLSTATE and YOOCHOOSE data, where we find that Algorithm 1 far outperforms its theoretical worst-case bound. We again considered all 2-item choice sets C and tested our method under MNL and CDM,[3] setting ε so that the experiment took about 30 minutes to run for ALLSTATE and 2 hours for YOOCHOOSE (of that time, Greedy takes 5 seconds to run; the rest is taken up by Algorithm 1). Algorithm 1 consistently outperforms Greedy (Fig. 2), even with $\varepsilon > 2$ for CDM. Moreover, Algorithm 1 only computes a small fraction of possible sets of alternatives, especially for YOOCHOOSE. Algorithm 1 does not perform as well with the rank-2 CDM as it does with MNL, which is to be expected as we only have approximation guarantees for CDM under particular parameter regimes (in which these data do not lie). The worse performance on CDM is due to the context effects that items from $\overline{C}$ exert on each other. Greedy does fairly well for DISAGREEMENT under CDM with YOOCHOOSE, but even in this case, Algorithm 1 performs significantly better in enough instances for the mean (but not median) performance to be better than Greedy. We repeated the experiment with 500 choice sets of size up to 5 sampled from data with similar results (Appendix E.3). We also ran the MIBLP approach for MNL, which performed as well as Algorithm 1 and was about 12x faster on YOOCHOOSE and 240x faster on ALLSTATE (Appendix E.2).

[3] In this case, we did not have available tree structures for NL, which are difficult to derive from data (Benson et al., 2016).

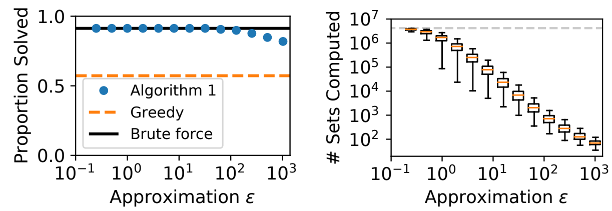

Figure 3. PROMOTION results on ALLSTATE 2-item choice sets. (Left) Success rate comparison; Algorithm 1 has near-optimal performance (about 9% of instances have no PROMOTION solution). (Right) Number of subsets of $\overline{C}$ computed by Algorithm 1 (dashed gray line at $2^{22} = 2^m$ for brute force computation).

PROMOTION. We applied the CDM PROMOTION version of Algorithm 1 to ALLSTATE, since this dataset is small enough to compute brute-force solutions. For each 2-item choice set C, we attempted to promote the less-popular item of the pair using brute-force, Greedy, and Algorithm 1. Algorithm 1 performed optimally up to $\varepsilon = 32$, above which it failed in only 2–26 of 252 feasible instances (Fig. 3, left). (Here, successful promotion means that the item becomes the true favorite among C.) On the other hand, Greedy failed in 37% of the feasible instances. As in the previous experiment, our algorithm's performance in practice far exceeds the worst-case bounds. The number of sets tested by Algorithm 1 falls dramatically as ε increases (Fig. 3, right). With more items (or a smaller range of utilities), the value of ε required to achieve the same speedup over brute force would be smaller (as with YOOCHOOSE). In tandem, these results show that we get near-optimal PROMOTION performance with far fewer computations than brute force.

7. Discussion

Our decisions are influenced by the alternatives that are available, the choice set. In collective decision-making, altering the choice set can encourage agreement or create new conflict. We formulated this as an algorithmic question: how can we optimize the choice set for some objective?

We showed that choice set optimization is NP-hard for natural objectives under standard choice models; however, we also found that model restrictions makes promoting a choice easier than encouraging a group to agree or disagree. We developed approximation algorithms for these hard problems that are effective in practice, although there remains a gap between theoretical approximation bounds and performance on real-world data. Future work could address choice set optimization in interactive group decisions, where group members can communicate their preferences to each other or must collaborate to reach a unified decision. Lastly, Appendix F discusses the ethical considerations of this work.

Acknowledgments

This research was supported by ARO MURI, ARO Award W911NF19-1-0057, NSF Award DMS-1830274, and JP Morgan Chase & Co. We thank Johan Ugander for helpful conversations.

References

Amir, O., Grosz, B. J., Gajos, K. Z., Swenson, S. M., and Sanders, L. M. From care plans to care coordination: Opportunities for computer support of teamwork in complex healthcare. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 1419–1428. ACM, 2015.

Ammar, A., Oh, S., Shah, D., and Voloch, L. F. What's your choice?: learning the mixed multi-nomial. In *ACM SIGMETRICS Performance Evaluation Review*, pp. 565–566. ACM, 2014.

Ben-Shimon, D., Tsikinovsky, A., Friedmann, M., Shapira, B., Rokach, L., and Hoerle, J. RecSys challenge 2015 and the YOOCHOOSE dataset. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pp. 357–358, 2015.

Benson, A. R., Kumar, R., and Tomkins, A. On the relevance of irrelevant alternatives. In *Proceedings of the 25th International Conference on World Wide Web*, pp. 963–973. International World Wide Web Conferences Steering Committee, 2016.

Benson, A. R., Kumar, R., and Tomkins, A. A discrete choice model for subset selection. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 37–45. ACM, 2018.

Bower, A. and Balzano, L. Preference modeling with context-dependent salient features. In *Proceedings of the 37th International Coference on International Conference on Machine Learning*, 2020.

Chen, S. and Joachims, T. Predicting matchups and preferences in context. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 775–784. ACM, 2016.

Chen, X., Lijffijt, J., and De Bie, T. Quantifying and minimizing risk of conflict in social networks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1197–1205. ACM, 2018.

Chierichetti, F., Kumar, R., and Tomkins, A. Discrete choice, permutations, and reconstruction. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 576–586. SIAM, 2018a.

Chierichetti, F., Kumar, R., and Tomkins, A. Learning a mixture of two multinomial logits. In *International Conference on Machine Learning*, pp. 960–968, 2018b.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. *Introduction to Algorithms*. MIT Press, 2001.

Davis, J. M., Gallego, G., and Topaloglu, H. Assortment optimization under variants of the nested logit model. *Operations Research*, 62(2):250–273, 2014.

Easley, D. and Kleinberg, J. *Networks, Crowds, and Markets*. Cambridge University Press, 2010.

Fader, P. S. and McAlister, L. An elimination by aspects model of consumer response to promotion calibrated on UPC scanner data. *Journal of Marketing Research*, 27(3): 322–332, 1990.

Gallego, G. and Topaloglu, H. Constrained assortment optimization for the nested logit model. *Management Science*, 60(10):2583–2601, 2014.

Garimella, K., De Francisci Morales, G., Gionis, A., and Mathioudakis, M. Reducing controversy by connecting opposing views. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 81–90. ACM, 2017.

Graells-Garrido, E., Lalmas, M., and Quercia, D. People of opposing views can share common interests. In *Proceedings of the 23rd International Conference on World Wide Web*, pp. 281–282. ACM, 2014.

Hanley, N., Wright, R. E., and Adamowicz, V. Using choice experiments to value the environment. *Environmental and Resource Economics*, 11(3-4):413–428, 1998.

Hązła, J., Jadbabaie, A., Mossel, E., and Rahimian, M. A. Bayesian decision making in groups is hard. *arXiv:1705.04770*, 2017.

Hązła, J., Jadbabaie, A., Mossel, E., and Rahimian, M. A. Reasoning in Bayesian opinion exchange networks is PSPACE-hard. In *Proceedings of the Thirty-Second Conference on Learning Theory*, pp. 1614–1648, 2019.

Huber, J., Payne, J. W., and Puto, C. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 9 (1):90–98, 1982.

Ieong, S., Mishra, N., and Sheffet, O. Predicting preference flips in commerce search. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, pp. 1795–1802. Omnipress, 2012.

Kaggle. Allstate purchase prediction challenge. https://www.kaggle.com/c/allstate-purchase-prediction-challenge, 2014.

Kallus, N. and Udell, M. Revealed preference at scale: Learning personalized preferences from assortment choices. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pp. 821–837. ACM, 2016.

Karp, R. M. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pp. 85–103. Springer, 1972.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Kleinberg, J., Mullainathan, S., and Ugander, J. Comparison-based choices. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pp. 127–144. ACM, 2017.

Koppelman, F. S. and Bhat, C. A self instructing course in mode choice modeling: Multinomial and nested logit models, 2006.

Liao, Q. V. and Fu, W.-T. Can you hear me now?: mitigating the echo chamber effect by source position indicators. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 184–196. ACM, 2014.

Lu, L., Yuan, Y. C., and McLeod, P. L. Twenty-five years of hidden profiles in group decision making: A meta-analysis. *Personality and Social Psychology Review*, 16 (1):54–75, 2012.

Luce, R. D. *Individual Choice Behavior*. Wiley, 1959.

Makhijani, R. and Ugander, J. Parametric models for intransitivity in pairwise rankings. In *The World Wide Web Conference*, pp. 3056–3062. ACM, 2019.

Matakos, A., Terzi, E., and Tsaparas, P. Measuring and moderating opinion polarization in social networks. *Data Mining and Knowledge Discovery*, 31(5):1480–1505, 2017.

Maystre, L. and Grossglauser, M. Fast and accurate inference of Plackett–Luce models. In *Advances in Neural Information Processing Systems*, pp. 172–180, 2015.

McFadden, D. Conditional logit analysis of qualitative choice behavior. In Zarembka, P. (ed.), *Frontiers in Econometrics*, pp. 105–142. Academic Press, 1974.

McFadden, D. Modeling the choice of residential location. *Transportation Research Record*, 1978.

McFadden, D. and Train, K. Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5): 447–470, 2000.

Munson, S. A., Lee, S. Y., and Resnick, P. Encouraging reading of diverse political viewpoints with a browser widget. In *Seventh International AAAI Conference on Weblogs and Social Media*, 2013.

Musco, C., Musco, C., and Tsourakakis, C. E. Minimizing polarization and disagreement in social networks. In *Proceedings of the 2018 World Wide Web Conference*, pp. 369–378. International World Wide Web Conferences Steering Committee, 2018.

Oh, S. and Shah, D. Learning mixed multinomial logit model from ordinal data. In *Advances in Neural Information Processing Systems*, pp. 595–603, 2014.

Oh, S., Thekumparampil, K. K., and Xu, J. Collaboratively learning preferences from ordinal data. In *Advances in Neural Information Processing Systems*, pp. 1909–1917, 2015.

Overgoor, J., Benson, A., and Ugander, J. Choosing to grow a graph: modeling network formation as discrete choice. In *The World Wide Web Conference*, pp. 1409–1420, 2019.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, pp. 8024–8035, 2019.

Pfannschmidt, K., Gupta, P., and Hüllermeier, E. Learning choice functions: Concepts and architectures. *arXiv:1901.10860*, 2019.

Ragain, S. and Ugander, J. Pairwise choice Markov chains. In *Advances in Neural Information Processing Systems*, pp. 3198–3206, 2016.

Reddi, S. J., Kale, S., and Kumar, S. On the convergence of Adam and beyond. In *International Conference on Learning Representations*, 2018.

Rosenberg, M., Perlroth, N., and Sanger, D. E. 'Chaos is the point': Russian hackers and trolls grow stealthier in 2020, 2020. URL https://www.nytimes.com/2020/01/10/us/politics/russia-hacking-disinformation-election.html.

Rosenfeld, N., Oshiba, K., and Singer, Y. Predicting choice with set-dependent aggregation. In *Proceedings of the 37th International Coference on International Conference on Machine Learning*, 2020.

Rusmevichientong, P., Shen, Z.-J. M., and Shmoys, D. B. Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations Research*, 58(6):1666–1680, 2010.

Rusmevichientong, P., Shmoys, D., Tong, C., and Topaloglu, H. Assortment optimization under the multinomial logit model with random choice parameters. *Production and Operations Management*, 23(11):2023–2039, 2014.

Seshadri, A. and Ugander, J. Fundamental limits of testing the independence of irrelevant alternatives in discrete choice. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pp. 65–66, 2019.

Seshadri, A., Peysakhovich, A., and Ugander, J. Discovering context effects from raw choice data. In *International Conference on Machine Learning*, pp. 5660–5669, 2019.

Shafir, E., Simonson, I., and Tversky, A. Reason-based choice. *Cognition*, 49(1-2):11–36, 1993.

Simonson, I. Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, 16(2):158–174, 1989.

Simonson, I. and Tversky, A. Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, 29(3):281–295, 1992.

Stasser, G. and Titus, W. Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, 48(6):1467, 1985.

Talluri, K. and Van Ryzin, G. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1):15–33, 2004.

Train, K. E. *Discrete Choice Methods with Simulation*. Cambridge University Press, 2009.

Trueblood, J. S., Brown, S. D., Heathcote, A., and Busemeyer, J. R. Not just for consumers: Context effects are fundamental to decision making. *Psychological Science*, 24(6):901–908, 2013.

Tversky, A. Elimination by aspects: A theory of choice. *Psychological Review*, 79(4):281, 1972.

Tversky, A. and Kahneman, D. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981.

Tversky, A. and Simonson, I. Context-dependent preferences. *Management Science*, 39(10):1179–1189, 1993.

Zhao, Z. and Xia, L. Learning mixtures of Plackett-Luce models from structured partial orders. In *Advances in Neural Information Processing Systems*, pp. 10143–10153, 2019.

Zhao, Z., Piech, P., and Xia, L. Learning mixtures of Plackett-Luce models. In *International Conference on Machine Learning*, pp. 2906–2914, 2016.