Whether manipulation is a problem in practice depends on: how likely such manipulable preference profiles are; whether manipulators can detect the existence such a profile; and whether computing a suitable misreport is feasible. On this third point, the pioneering work of Bartholdi et al. [2, 3] demonstrated that, even given full knowledge of a profile, computing a suitable manipulation is computationally intractable for certain voting rules. This in turn led to the detailed computational analysis of many voting rules (e.g., the Borda rule [13, 4]). Of course, worst-case complexity results offer little comfort if “difficult profiles” are unlikely to arise in practice. Recent work suggests that common voting rules are in fact frequently manipulable, by studying heuristic algorithms that provide theoretical guarantees [31, 41, 39], identifying properties of voting rules that make them easy to manipulate in the typical case [11, 16, 40], and investigating (both theoretically and empirically) the relation between the number of manipulators and the probability of manipulation [30, 38, 35] (see [15] for an overview).

Analyses demonstrating ease of manipulation tend to suffer from two key drawbacks. First, they exclusively analyze manipulation assuming the manipulating coalition has full knowledge of the vote profile. While results showing that manipulation is difficult can be justified on these grounds, claiming easiness of manipulation has less practical import if the coalition is assumed to have unreasonable access to the preferences of sincere voters.\(^3\) One exception considers manipulators who know only that the vote profile lies within some set [12], but unfortunately this work only analyzes the rather weak notion of dominating manipulations. Social choice research on manipulation under probabilistic knowledge is mostly negative in nature [23], or restricted to a single manipulator [1].

A second weakness of many analyses of probability of manipulation (which do assume complete information on the part of the manipulator) is their reliance on specific stylized models such as impartial culture (where every ranking in equally likely) [16, 40]. Much empirical work also considers very stylized distributions such as impartial culture, Polya’s urn, and Condorcet (or Mallows) distributions. Some work does consider sub-sampling from real voting data, though with relatively small numbers of votes [36].

### 2.3 Probabilistic Ranking Models

Probabilistic analysis of manipulation—including our Bayesian manipulation problem—requires some probabilistic model of voter preferences. By far the most common model in social choice is impartial culture (IC), which assumes the preference of any voter is drawn from the uniform distribution over the set of permutations of alternatives [32]. A related model is the impartial anonymous culture (IAC) model in which each voting situation is equally likely [32].\(^4\) Several other models (bipolar, urn, etc.) are considered in both theoretical and empirical social choice research.

Probabilistic models of rankings are widely considered in statistics, econometrics and machine learning as well, including models such as Mallows φ-model, Plackett-Luce, and mixtures thereof [25]. We use the Mallows φ-model [24] in Sec. 6, which is parameterized by a reference ranking σ and a dispersion φ ∈ (0,1], with

\[ P(r) = \frac{1}{Z} \phi^{d(r, σ)}, \]

where \( r \) is any ranking, \( d \) is Kendall’s τ-distance, and \( Z \) is a normalizing constant. When \( φ = 1 \), this model is exactly the impartial culture model studied widely in social choice—as such it offers considerable modelling flexibility. However, mixtures of Mallows models offer even greater flexibility, allowing (with enough mixture components) accurate modelling of any distribution over preferences. As a consequence, Mallows models, and mixtures thereof, have attracted considerable attention in the machine learning community [27, 8, 20, 26, 21]. We investigate these models empirically below.

### 3 Optimal Bayesian Manipulation

We now consider how a manipulating coalition should act given probabilistic knowledge of the preferences of the sincere voters. We first formally define our setting, then present several analytical results. Finally, we present a general, sample-based optimization framework for computing optimal manipulation strategies and provide sample complexity results for positional scoring rules and k-approval.

#### 3.1 The Model

We make the standard assumption that voters are partitioned into \( n \) sincere voters, who provide their true rankings to a voting mechanism or rule \( r \), and a coalition of \( c \) manipulators. We assume the manipulators have a desired alternative \( d \in A \), and w.l.o.g. we assume \( A = \{a_1, \ldots, a_m = d\} \). We make no assumptions about the manipulators’ specific preferences, only that they desire to cast their votes so as to maximize the probability of \( d \) winning under \( r \). A vote profile can be partitioned as \( v = (v_n, v_c) \), where \( v_n \) reflects the true

---

\(^3\)This is implicit in [10], which shows that hardness of full-information manipulation implies hardness under probabilistic information.

\(^4\)A voting situation simply counts the number of voters who hold each possible ranking of the alternatives.