On Determining Dodgson Winners by Frequently Self-Knowingly Correct Algorithms and in Average-Case Polynomial Time\footnote{Supported in part by the German Science Foundation under grants RO 1202/9-1 and RO 1202/9-3.}

Jörg Rothe\footnote{Supported in part by the Alexander von Humboldt Foundation in the TransCoop program.} and Holger Spakowski

Abstract

In their study of an efficient greedy heuristic for determining Dodgson winners, Homan and Hemaspaandra [HH06] introduced the notion of frequently self-knowingly correct algorithm. We show that this notion is closely related to Impagliazzo's notion of polynomial-time benign algorithm scheme [Imp95], which provides a model of average-case polynomial time. In particular, we show that every distributional problem (with respect to the uniform distribution) that has a polynomial-time benign algorithm scheme also has a frequently self-knowingly correct algorithm. Furthermore, we discuss Homan and Hemaspaandra's greedy heuristic with respect to AvgP, Levin's notion of average polynomial time [Lev86].

Key words: Preference aggregation, election systems, Dodgson elections, frequently self-knowingly correct algorithms, average-case complexity.

1 Introduction

This paper studies a novel type of algorithm, called frequently self-knowingly correct algorithm, and its relation to average-case polynomial time. Frequently self-knowingly correct algorithms were introduced by Homan and Hemaspaandra [HH06], who designed efficient such algorithms for solving the winner problem for Dodgson elections (which is known to be hard in the worst-case complexity model) with a guaranteed high frequency of success.

Before we turn to the main purpose of this paper, let us give a brief overview of recent results on complexity-theoretic issues related to voting in order to motivate this paper's topic in a broader framework. For more background on computational politics and the complexity of electoral problems, we refer to the comprehensive surveys [FHHR06, HH00].

Preference aggregation and election systems have been studied for centuries in social choice theory, political science, and economics, see, e.g., Arrow [Arr63],...
Black [Bla58], Brams and Fishburn [BF83], and McLean and Urken [MU95]. Recently, these topics have become the focus of attention in various areas of computer science as well, such as artificial intelligence (especially with regard to distributed AI in multi-agent settings), computational complexity, and operations research. In the field of computational complexity, much work has been done during the past few years on the following four important classes of problems for various election systems.

Let $\mathcal{E}$ be a given election system. The **winner problem** for $\mathcal{E}$ asks whether a designated candidate has won a given election under $\mathcal{E}$. Bartholdi, Tovey, and Trick [BTT89b] proved that the winner problem for both Dodgson elections [Dod76] and Kemeny elections [Kem59, KS60] is NP-hard. Hemaspaandra, Hemaspaandra, and Rothe [HHR97] optimally improved the former result by proving the Dodgson winner problem complete (under polynomial-time many-one reductions) for $P^{NP}$, the class of problems solvable via parallel access to NP. This class is known to be equal to a number of other classes (including $P_{\text{truthful}}^{NP}$ and $P_{\text{NPlog}}^{NP}$, see [PZ83, Wag90]) and forms the $\Theta_2^p$ level of the polynomial hierarchy.

Rothe, Spakowski, and Vogel [RSV03] proved that the winner problem for Young elections [You77] is also $P^{NP}$-complete, and Hemaspaandra, Spakowski, and Vogel [HSV05] obtained the analogous result for the Kemeny winner problem. Each of these three election systems is based on a certain combinatorial optimization problem, and while each of these systems avoids the Condorcet Paradox, it does respect the Condorcet Principle [Con85, Fis77], i.e., it selects the Condorcet winner whenever one exists.¹ All the above results focus on “winner problems” (which ask if the designated candidate wins regardless of whether there are other winners as well) as opposed to “unique winner problems,” and we also adopt the traditional model of winner problem here.²

The **control problem** for election system $\mathcal{E}$ asks whether it is feasible for an election chair to alter the outcome of an election by changing its agenda (see, e.g., [BTT92, HHR05]). The **manipulation problem** (a.k.a. the *strategic voting problem*) for $\mathcal{E}$ asks whether it is feasible that the outcome of an election can be altered by having voters strategically change their preferences (see, e.g., [BTT89a, BO91, CS02, CLS03, HH05]). The **bribery problem** for $\mathcal{E}$ (which is somewhat related to manipulation) asks whether it is feasible that the outcome of an election can be altered by an agent who bribes voters to change their votes. Of course, most desirable are voting systems whose winner problem is easy yet which resist electoral control, manipulation, and bribery.

We are concerned with the winner problem only. Since for some election systems with otherwise useful properties the winner problem is hard (in the worst-case complexity model), it is natural to wonder if one at least can find

---

¹A Condorcet winner is a candidate who beats all other candidates in pairwise majority-rule contests.

²Note, however, that Hemaspaandra, Hemaspaandra, and Rothe [HHR06] have recently shown that also the unique winner problems for Dodgson, Young, and Kemeny elections are $P_1^{NP}$-complete.
a heuristic algorithm solving the problem for “most of the inputs occurring in practice.” Examples of such heuristics are known, e.g., for the Dodgson winner problem [HH06] and for the web page ranking problem in close relation to the Kemeny winner problem [DKNS01]. In particular, we study a heuristic called GreedyWinner for the problem of determining the winners of Dodgson elections. This heuristic is due to Homan and Hemaspaandra [HH06], who proved that if the number of voters greatly exceeds the number of candidates (which in many real-world cases is a very plausible assumption), then their heuristic is a “frequently self-knowingly correct algorithm,” a notion they introduced in [HH06]. We show that this notion is closely related to average-case complexity.

This paper is organized as follows. In Section 2, we give a brief introduction to social choice theory (and, in particular, to Dodgson elections), we present some foundations of average-case complexity theory, and we define the notion of frequently self-knowingly correct algorithm. Section 3 provides our main result: Every problem in AvgP has a frequently self-knowingly correct algorithm. In Section 4, we discuss the relation of Homan and Hemaspaandra’s greedy heuristic for finding Dodgson winners to average-case polynomial time. Finally, in Section 5, we conclude by raising some related open questions.

2 Preliminaries

2.1 Some Background from Social Choice Theory

An election $E = (C, V)$ is given by a set $C$ of candidates and a set $V$ of votes, where each vote is specified by a preference order on all candidates. As is common, we assume that the underlying preference relation is strict (i.e., irreflexive and asymmetric), transitive, and complete.

In this paper, we focus on Dodgson elections. In 1876, Dodgson proposed an election system [Dod76] that is based on a combinatorial optimization problem: An election is won by those candidates who are “closest” to being a Condorcet winner, the unique candidate (if one exists) who defeats each other candidate in pairwise comparison by a strict majority of votes.

More precisely, given a Dodgson election $E = (C, V)$, every candidate $c$ in $C$ is assigned a score, which is denoted by $\text{DodgsonScore}(C, V, c)$ and is defined to be the smallest number of sequential exchanges of adjacent preferences in the voters’ preference orders needed to make $c$ a Condorcet winner with respect to the resulting preference orders. Whoever has the lowest Dodgson score wins.

The problem $\text{DodgsonWinner}$ is defined as follows: Given an election $E = (C, V)$ and a designated candidate $c$ in $C$, is $c$ a Dodgson winner in $E$? (The search version of this decision problem can easily be derived.) As mentioned above, Hemaspaandra, Hemaspaandra, and Rothe [HHR97] have shown that this problem is $\text{P}^\text{NP}$-complete.

It certainly is not desirable to have an election system whose winner problem is hard, as only systems that can be evaluated efficiently are actually used in practice. Fortunately, there are a number of positive results on Dodgson
elections and related systems as well. In particular, Bartholdi, Tovey, and Trick [BTT89b] proved that for elections with a bounded number of candidates or voters Dodgson winners are easy to determine. Fishburn [Fis77] proposed a “homogeneous” variant of Dodgson elections that Rote, Spakowski, and Vogel [RSV03] proved to have a polynomial-time winner problem. McCabe-Dansted, Pritchard, and Slinko [MDPS06] proposed a scheme (called Dodgson Quick) that approximates Dodgson’s rule with an exponentially fast convergence. Finally, Homan and Hemaspaandra [HH06] proposed a greedy heuristic that finds Dodgson winners with a guaranteed high frequency of success. In particular, they introduced the notion of “frequently self-knowingly correct algorithm,” which we define in Section 2.3 below, and they noted:

“The closest related concepts […] are probably those involving proofs to be verified, such as NP certificates and the proofs in interactive proof systems.”

This statement notwithstanding, we will show that in fact it is the theory of average-case complexity (which Homan and Hemaspaandra also mention in their paper) that is even more closely related to their notion.

2.2 Foundations of Average-Case Complexity Theory

The theory of average-case complexity was initiated by Levin [Lev86]. A problem’s average-case complexity can be viewed a more significant measure than its worst-case complexity in many cases, for example in cryptographic applications. We here follow Goldreich’s presentation [Go97]. Another excellent introduction to this theory is due to Wang [Wan97].

Fix the alphabet \( \Sigma = \{0, 1\} \), let \( \Sigma^* \) denote the set of strings over \( \Sigma \), and let \( \Sigma^n \) denote the set of all length \( n \) strings in \( \Sigma^* \). For any \( x, y \in \Sigma^* \), \( x < y \) means that \( x \) precedes \( y \) in lexicographic order, and \( x - 1 \) denotes the lexicographic predecessor of \( x \).

Intuitively, Levin observed that many hard problems—including those that are NP-hard in the traditional worst-case complexity model—may nonetheless be easy to solve “on the average,” i.e., for “most” inputs or for “most practically relevant” inputs. He proposed to define the complexity of problems with respect to some suitable distribution on the input strings.

We now define the notion of a distributional problem and the complexity class \( \text{AvgP} \). In subsequent sections, we consider two heuristic algorithms: the algorithm \text{GreedyWinner} intended to solve the decision problem \text{DodgsonWinner}, and the algorithm \text{GreedyScore} intended to compute the Dodgson score of some given candidate. Both heuristics work well sufficiently often, provided that the number of voters greatly exceeds the number of candidates.

Here, we define only distributional search problems; the definition of distributional decision problems is analogous.
Definition 1  
1. A distribution function \( \mu : \Sigma^* \rightarrow [0, 1] \) is a nondecreasing function from strings to the unit interval that converges to one (i.e., \( \mu(0) \geq 0 \) and \( \mu(x) \leq \mu(y) \) for each \( x < y \), and \( \lim_{x \rightarrow \infty} \mu(x) = 1 \)). The density function associated with \( \mu \) is defined by \( \mu'(0) = \mu(0) \) and \( \mu'(x) = \mu(x) - \mu(x - 1) \) for each \( x > 0 \). That is, each string \( x \) gets weight \( \mu'(x) \) with this distribution.

2. A distributional (search) problem is a pair \((f, \mu)\), where \( f : \Sigma^* \rightarrow \Sigma^* \) is a function and \( \mu : \Sigma^* \rightarrow [0,1] \) is a distribution function.

3. A function \( t : \Sigma^* \rightarrow \mathbb{N} \) is polynomial on the average with respect to some distribution \( \mu \) if there exists a constant \( \epsilon > 0 \) such that
\[
\sum_{x \in \Sigma^n} \mu'(x) \cdot \frac{t(x)^\epsilon}{|x|} < \infty.
\]

4. The class \( \text{AvgP} \) consists of all distributional problems \((f, \mu)\) for which there exists an algorithm \( A \) computing \( f \) such that the running time of \( A \) is polynomial on the average with respect to the distribution \( \mu \).

In this paper, we focus on the standard uniform distribution \( \mu \) on \( \Sigma^* \), which is defined by
\[
\mu'(x) = \frac{1}{|x|(|x| + 1)^{|x|}}.
\]
That is, we first choose an input size \( n \) at random with probability \( 1/n(n+1) \), and then we choose an input string of that size \( n \) uniformly at random. For each \( n \in \mathbb{N} \), let \( \mu_n \) be the restriction of \( \mu \) to strings of length at most \( n \).

Impagliazzo [Imp95] introduced the notion of polynomial-time benign algorithm scheme to present an alternative view on the definition of average polynomial time.

Definition 2 ([Imp95])  
1. An algorithm computes a function \( f \) with benign faults if it either outputs an element of the image of \( f \) or "?", and if it outputs anything other than \?, it is correct.

2. A polynomial-time benign algorithm scheme for a function \( f \) on \( \mu_n \) is an algorithm \( A(x, \delta) \) such that:
   (a) \( A \) runs in time polynomial in \( |x| \) and \( 1/\delta \).
   (b) \( A \) computes \( f \) with benign faults.
   (c) For each \( \delta, 0 < \delta < 1 \), and for each \( n \in \mathbb{N}^* \), \( \Pr_{\mu_n}[A(x, \delta) =?] \leq \delta \).
2.3 A Frequently Self-Knowingly Correct Greedy Algorithm

Homan and Hemaspaandra [HH06] define the following notion.

**Definition 3 ([HH06])** 1. Let $f : S \rightarrow T$ be a function, where $S$ and $T$ are sets. We say an algorithm $A : S \rightarrow T \times \{ \text{"definitely", \ "maybe"} \}$ is self-knowingly correct for $f$ if, for each $s \in S$ and $t \in T$, whenever $A$ on input $s$ outputs $(t, \text{"definitely"})$ then $f(s) = t$.

2. An algorithm $A$ that is self-knowingly correct for $g : \Sigma^* \rightarrow T$ is said to be frequently self-knowingly correct for $g$ if

$$\lim_{n \to \infty} \frac{||\{x \in \Sigma^n | A(x) \in T \times \{ \text{"maybe"} \}\}||}{||\Sigma^n||} = 0.$$ 

In their seminal paper [HH06], Homan and Hemaspaandra present two frequently self-knowingly correct polynomial-time algorithms, which they call GreedyScore and GreedyWinner. Since GreedyWinner can easily be reduced to GreedyScore, we focus on GreedyScore only and briefly describe the intuition behind this algorithm; for full detail, we refer to [HH06].

If $E = (C, V)$ is an election and $c$ is some designated candidate in $C$, we call $(C, V, c)$ a Dodgson triple. Given a Dodgson triple $(C, V, c)$, GreedyScore determines the Dodgson score of $c$ with respect to the given election $(C, V)$. We will see that there are Dodgson triples $(C, V, c)$ for which this problem is particularly easy to solve.

For any $d \in C - \{c\}$, let $\text{Deficit}[d]$ be the number of votes $c$ needs to gain in order to have more votes than $d$ in a pairwise contest between $c$ and $d$.

**Definition 4** Any Dodgson triple $(C, V, c)$ is said to be nice if for each candidate $d \in C - \{c\}$, there are at least $\text{Deficit}[d]$ votes for which candidate $c$ is exactly one position below candidate $d$.

Given a Dodgson triple $(C, V, c)$, the algorithm GreedyScore works as follows:

1. For each candidate $d \in C - \{c\}$, determine $\text{Deficit}[d]$.

2. If $(C, V, c)$ is not nice then output ("anything", "maybe"); otherwise, output

$$\sum_{d \in C - \{c\}} \text{Deficit}[d], \text{"definitely"}.$$ 

Note that, for nice Dodgson triples, we have

$$\text{DodgsonScore}(C, V, c) = \sum_{d \in C - \{c\}} \text{Deficit}[d],$$ 

It is easy to see that GreedyScore is a self-knowingly correct polynomial-time bounded algorithm. To show that it is even frequently self-knowingly correct, Homan and Hemaspaandra prove the following key lemma. Their proof uses a variant of Chernoff bounds.
**Lemma 5 (see Thm. 4.1(3) in [HH06])** Let $(C, V, c)$ be a given Dodson triple with $n = ||V||$ votes and $m = ||C||$ candidates, chosen uniformly at random among all such Dodson elections. The probability that $(C, V, c)$ is not nice is at most

$$2(m - 1)e^{-\frac{m^2}{8n}}.$$

Homan and Hemaspaandra [HH06] show that the heuristic GreedyWinner, which is based on GreedyScore and which solves the winner problem for Dodson elections, also is a frequently self-knowingly correct polynomial-time algorithm. This result is stated formally below.

**Theorem 6 (HH06)** For all $m, n \in \mathbb{N}^+$, the probability that a Dodson election $E$ selected uniformly at random from all Dodson elections having $m$ candidates and $n$ votes (i.e., all $(m!)^n$ Dodson elections having $m$ candidates and $n$ votes have the same likelihood of being selected) has the property that there exists at least one candidate $c$ such that GreedyWinner on input $(E, c)$ outputs “maybe” as its second output component is less than $2(m^2 - m)e^{-\frac{m^2}{8n}}$.

### 3 On AvgP and Frequently Self-Knowingly Correct Algorithms

Our main result relates polynomial-time benign algorithm schemes to frequently self-knowingly correct algorithms. We show that every distributional problem (with respect to the uniform distribution) that has a polynomial-time benign algorithm scheme must also have a frequently self-knowingly correct algorithm. It follows that all AvgP problems have a frequently self-knowingly correct algorithm.

**Theorem 7** Suppose that $\mathcal{A}(x, \delta)$ is a polynomial-time benign algorithm scheme for a distributional problem $f$ on $\mu_n$. Then there is a frequently self-knowingly correct algorithm $\mathcal{A}'$ for $f$.

**Proof** For each $n \in \mathbb{N}$, let $\delta(n) = 1/n^3$. Define the algorithm $\mathcal{A}'$ as follows:

1. On input $x \in \Sigma^*$, simulate $\mathcal{A}(x, \delta(|x|))$.
2. If $\mathcal{A}(x, \delta(|x|))$ outputs $?$, then output (anything, “maybe”).
3. If $\mathcal{A}(x, \delta(|x|))$ outputs $y \in T$, where $y \neq ?$, then output $y$, “definitely”.

By definition of “polynomial-time benign algorithm scheme,” algorithm $\mathcal{A}'$ is polynomial-time bounded. It remains to show that $\mathcal{A}'$ is frequently self-knowingly correct.
Fix an arbitrary \( n \in \mathbb{N} \). Then

\[
\frac{||\{x \in \Sigma^n \mid A'(x) \in T \times \{"maybe"\}\}||}{||\Sigma^n||} = \frac{\text{Prob}_{\mu_n}[A(x, \delta(|x|)) = ? | |x| = n]}{\text{Prob}_{\mu_n}[|x| = n]} \\
\leq \frac{\text{Prob}_{\mu_n}[A(x, \delta(|x|)) = ?]}{1/n(n+1)} \\
= \frac{n(n+1) \cdot \text{Prob}_{\mu_n}[A(x, \delta(|x|)) = ?]}{n(n+1) \cdot \delta(|x|)} \\
\leq \frac{n+1}{n^2}.
\]

Here, (1) holds because, by definition of \( \mu_n \),

\[\text{Prob}_{\mu_n}[|x| = n] \geq 1/n(n+1),\]

and (2) is true by the definition of polynomial-time benign algorithm scheme.

We have shown that

\[\lim_{n \to \infty} \frac{||\{x \in \Sigma^n \mid A'(x) \in T \times \{"maybe"\}\}||}{||\Sigma^n||} = 0,\]

which completes the proof.

**Corollary 8** Every distributional problem (under the standard uniform distribution) that is in AvgP has a frequently self-knowingly correct algorithm.

**Proof** Impagliazzo proved that any distributional problem on input ensemble \( \mu_n \) is in AvgP if and only if it has a polynomial-time benign algorithm scheme; see Proposition 2 in [Imp95]. The claim now follows from Theorem 7.

\[\square\]

## 4 Dodgson Winners and Average-Case Polynomial Time

Because of the close relationship between the notion of frequently self-knowingly correctness and average-case complexity, one might think that Homan and Hemaspaandra’s algorithm could be used to devise an algorithm, call it
DodgsonWinner-AverageGood, witnessing that the problem DodgsonWinner is in AvgP (assuming the uniform probability distribution on all elections with the same number of candidates and voters). A tempting approach towards this goal would be as follows. Given an election, run the GreedyWinner algorithm on it. If that fails (i.e., if GreedyWinner outputs “maybe”), then use the exhaustive algorithm that works in time \( O(m^n) \).

We now show that this approach does not work in any obvious way. The reason is that the algorithm is not “frequently enough” self-knowingly correct. That is, DodgsonWinner-AverageGood would have to spend too much time (namely, \( O(m^n) \)) on a too large fraction of the inputs, namely,

\[
2(m^2 - m)e^{-\frac{\epsilon}{2\pi r^2}}.
\]

Let \( \mu'(m, n) \) be the probability that an election chosen randomly under the uniform distribution has \( m \) candidates and \( n \) voters. We assume that the elections are encoded into strings in a reasonable way. Use, for instance, the encoding scheme given by Homan and Henaspaandra [HH06].

To prove that DodgsonWinner is in AvgP, we would have to show that there is an \( \epsilon > 0 \) such that

\[
\sum_{x \in \Sigma^*} \mu'(x) \frac{t(x)^\epsilon}{|x|} < \infty,
\]

where \( t(x) \) is the computation time of the algorithm DodgsonWinner-AverageGood on input \( x \), see Definition 1.

Let \( E(m, n) \) be the set of strings in \( \Sigma^* \) that encode elections having \( m \) candidates and \( n \) voters.

For any \( x \in E(m, n) \),

\[
\mu'(x) = \frac{1}{|E(m, n)|} \cdot \mu'(m, n).
\]

By Theorem 6, for fixed \( m \) and \( n \), we obtain

\[
\sum_{x \in E(m, n)} \mu'(x) \frac{t(x)^\epsilon}{|x|} \leq \mu'(m, n) \cdot \left( \frac{p(|x|)^\epsilon}{|x|} + \alpha(m, n) \cdot (cm^n)^\epsilon \right),
\]

where \( \alpha(m, n) = 2(m^2 - m)e^{-\frac{\epsilon}{2\pi r^2}} \). Here, \( p \) is the polynomial that bounds the computation time of GreedyWinner. We can clearly choose \( \epsilon > 0 \) small enough such that \( p(|x|)^\epsilon/|x| \) contributes only a constant in the above term. However, the crucial fact is that the term

\[
\alpha(m, n) \cdot (cm^n)^\epsilon = 2(m^2 - m)e^{-\frac{\epsilon}{2\pi r^2}} \cdot (cm^n)^\epsilon
\]

grows exponentially, no matter how small we choose \( \epsilon \).

Thus, DodgsonWinner-AverageGood does not run in average-case polynomial time with respect to any interesting distribution on the inputs.
5 Conclusions

Homan and Hemaspaandra [HH06] proposed a greedy heuristic for finding Dodgson winners, and they proved that this heuristic is a frequently self-knowingly correct algorithm. We have shown that every distributional problem (with respect to the standard uniform distribution) in $\text{AvgP}$ has a frequently self-knowingly correct algorithm. It is easy to see that the converse implication is not true. (For instance, one can define a problem $L$ that consists only of strings in $\{0\}^*$ encoding the halting problem. This problem is clearly not in $\text{AvgP}$, yet it is frequently self-knowingly correct.)

Furthermore, we have argued why it might be hard to show that the problem of finding Dodgson winners—at least via Homan and Hemaspaandra’s heuristic—can be solved in polynomial time on the average. We suspect that this problem is hard on the average, but a rigorous proof for this claim has eluded us so far, and we raise this as an open question. In particular, it would be interesting to study the average-case complexity of the Dodgson winner problem with respect to suitable distributions (see, e.g., Procaccia and Rosen-schein [PR06]).

Other interesting issues remain open as well. For example, one might study the approximability of the Dodgson winner problem; some first steps in this direction have been taken by McCabe-Dansted, Pritchard, and Slinko [MDPS06]. Also, one could investigate other voting systems with a hard winner problem in the worst-case model, such as the problem of determining Young winners [RSV03] or the problem of determining Kennedy winners [HSV05], and seek to find algorithms that can be shown to be frequently self-knowingly correct or even average-case polynomial-time, or else seek to prove these problems hard on the average.

Acknowledgments: We thank Lane A. Hemaspaandra and Chris Homan for their interest in this work, and for many inspiring discussions on computational issues related to voting. We also thank the COMSOC ’06 referees for their helpful comments.

References


[Dod76] C. Dodgson. A method of taking votes on more than two issues. Pamphlet printed by the Clarendon Press, Oxford, and headed “not yet published” (see the discussions in [MU95, Bla58], both of which reprint this paper), 1876.


Jörg Rothe
Institut für Informatik
Heinrich-Heine-Universität Düsseldorf
40225 Düsseldorf, Germany
Email: rothe@cs.uni-duesseldorf.de

Holger Spakowski
Institut für Informatik
Heinrich-Heine-Universität Düsseldorf
40225 Düsseldorf, Germany
Email: spakovsk@cs.uni-duesseldorf.de