Properties of the Win Probability in Penney's Ante

Mathew Drexel, April Ju, Peter Peng, and Jacob Richey

September 15, 2024

Abstract

Fix two words over the binary alphabet $\{0, 1\}$, and generate iid Bernoulli(p) bits until one of the words occurs in sequence. This game, commonly known as Penney's ante, was popularized by Conway, who found a simple formula for the probability that one of the two words wins the race. We study the win probability from an analytic and combinatorial perspective, in terms of the parameters v, w and p. For v and w of arbitrary lengths, our results bound how large the win probability can be for the longer word. When $p = \frac{1}{2}$ we obtain a characterization of the optimum, and for $p \neq \frac{1}{2}$ we present a conjecture along similar lines, which is supported by computer computations. For a fixed pair of words the win probability, viewed as a function of p, often exhibits nice properties, such as symmetry under the transformation $p \to 1 - p$. We find new explicit bijections that account for these symmetries, under conditions that can be easily verified by examining auto- and cross-correlations of the words.

1 Introduction

Penney's ante refers to the following random 'race': pick two finite words $v = v_1 \cdots v_m$ and $w = w_1 \cdots w_n$ over the alphabet $\{0, 1\}$, and generate an iid Bernoulli(p) sequence $(X_n)_{n \geq 1}$. The race ends when either v or w occurs in X, i.e. at time $\min\{\tau_v, \tau_w\}$, where for any binary word $u = u_1 \cdots u_r$, τ_u is the stopping time

$$\tau_u = \min\{t \in \mathbb{N} : X_{t-r+1} = u_1, X_{t-r+2} = u_2, \cdots X_t = u_r\}. \tag{1.1}$$

If word v occurred first, i.e. if $\tau_v < \tau_w$, then v wins the race. Assuming that neither v nor w is a subword of the other, the race is non-trivial in that both words win with positive probability. A basic study of the probability distributions τ_v appears in Feller's classic probability text [2]. Conway popularized the 'race' version as an example of a non-transitive tournament, and gave an explicit formula for the odds that one word occurs before another [4]. The formula, summarized in Theorem 4, arises from the so-called 'ABRACADABRA' martingale construction, which is well-known to probabilists [7]. It was generalized by Guibas/Odlyzsko and Li, who found linear systems describing the win probabilities and relevant generating functions for a race among an arbitrary (finite) family of words [4].

Our aim in this work is to give a birds eye view of the phenomena arising from the win probability $\mathbb{P}_p(\tau_v < \tau_w)$, where $\mathbb{P}_p$ denotes iid Bernoulli(p) measure. We stick to the iid case, and note that results similar to ours could likely be obtained for digits generated from an arbitrary finite-state markov chain [6][3]. The overarching questions we have in mind are: for fixed $p \in (0, 1)$, can we describe the optimum win probability, and the optimizing pairs (v, w)? Under what conditions on the words v, w does the win probability, viewed as a function of p, have some special property, like monotonicity or symmetry? When can those symmetries be explained by a bijection? While the first question has many papers about it, there is less work on the underlying combinatorial structure in this general setting.

For fixed $p \in (0,1)$, we study which pairs v, w give the largest possible win probability. Previous works identified the optimum for words of the same length at $p = \frac{1}{2}$ [8][1]: we extend this to words of different lengths. Our findings are in agreement with a well-known rule of thumb for how to choose a word v to beat a given word w [4]. We confirm in a detailed way that the rule holds for words of arbitrary length at $p = \frac{1}{2}$, and prove that the optimal pair occurs when the shorter word has period 1, i.e. is all 0s or all 1s. In the case $p \neq \frac{1}{2}$, we describe a conjecture along similar lines, which is supported by computer computations. Interestingly, in this case the optimum can occur for words that are essentially periodic with period 1 or 2, depending on the value of p. We do not fully understand this phenomenon – see section 6 for details.

One naturally expects that if p is perturbed away from $\frac{1}{2}$ – decreased, say – then it becomes a more common occurrence for a longer word to beat a shorter one. For p close to 0, we show in Theorem 9 that the winner is essentially determined by which word has fewer 1s, disregarding the total length of the word (though among words with the same number of 1s, there is a slight subtlety). Computer computations have suggested the following:

Conjecture 1. *For sufficiently large integer n, the number of pairs of binary words (v, w) of length at most n such that v is longer than w and $\mathbb{P}_p(\tau_v < \tau_w) > \frac{1}{2}$ is non-increasing for $p \in \left(0, \frac{1}{2}\right)$.*

A natural guess for how to prove Conjecture 1 is to show that for each fixed pair (v, w) of words, the win probability $\mathbb{P}_p(\tau_v < \tau_w)$ is a monotone (or constant) function of $p \in \left(0, \frac{1}{2}\right)$. Perhaps surprisingly, this is false: a simple example is the pair $(100010, 001100)$, where the minimum win probability is $\frac{2869}{6250}$, which occurs for $p = \frac{4}{5}$.

For certain pairs v, w, the win probability can exhibit different kinds of symmetry under the transformation $p \to 1 - p$, i.e. exchanging 0s and 1s. There are some obvious ways this can occur: as an example, if the word w is the bit-flip of the word v, then exchanging all 0s and 1s in both the sequence X and the words v, w shows that $\mathbb{P}_p(\tau_v < \tau_w) = \mathbb{P}_{1-p}(\tau_w < \tau_v)$. In the language of probability, flipping all the bits gives a coupling between Penney's ante played between the pair v, w and the pair $\overline{v}, \overline{w}$, the bit-flips of v and w, such that v wins the v, w race if and only if $\overline{w}$ wins the $\overline{v}, \overline{w}$ race.

Of course, using Equation 2.4, the win probability is always a rational function of p, and one can explicitly check whether some symmetry holds. However, we find that many symmetries like this occur for non-obvious reasons. This leads us to study the underlying combinatorial structure directly: namely, we consider the set Ω of all finite $\{0,1\}$ sequences that avoid the words v and w, and try to find explicit bijections between certain subsets of this set. When two sets are known to have the same cardinality, often from a recursion or generating function argument, it is common to seek an explicit bijection proof. In our setting, where the underlying iid sequence is Bernoulli(p), the Conway formula (or some graph-counting recursions or generating functions) guarantees that some two events $A, B \subset \Omega$ have the same probability under $\mathbb{P}_p$, and we seek a sufficiently nice bijection $\varphi : A \to B$. Here 'nice' means φ should transform the measure $\mathbb{P}_p$ according to the symmetry, e.g. in our bit-flip example $\mathbb{P}_p(\varphi(a)) = \mathbb{P}_{1-p}(a)$; and possibly that φ should be locally computable, i.e. to determine a given digit of $\varphi(a)$, one should only have to check a bounded-sized window of a, or a window of finite expected size. We present families of pairs where we have such explicit bijections in Proposition 10 and Theorem 12. We also find an explicit bijection for the pair $v = 1100, w = 1010$ which arises from a natural graph construction, and demonstrates the equality $\mathbb{P}_p(\tau_v < \tau_w) = \mathbb{P}_{1-p}(\tau_v < \tau_w)$.

Remark: In symbolic dynamics, a field at the intersection of combinatorics, dynamics, and ergodic theory, the sets Ω, called 'shift spaces,' and the maps φ between them, called 'conjugacies,' are commonplace objects of study. We note that the condition appearing in Theorem 10 gives a class of simple conjugacies, while the more interesting bijections constructed in the proof of Lemma 20 are not conjugacies. Typically in symbolic dynamics one studies the 'measure of maximal entropy' for a shift space – here we work with the measures $\mathbb{P}_p$, which are never maximal entropy measures for the corresponding shift spaces.

1.1 Overview

This work is structured as follows. Section 2 introduces the notation and formulas involved, most important among them being the Conway formula, Equation 2.4, for the win probability. In Section 3 we briefly advertise our main results and their implications. Section 4 covers the results and conjectures regarding words of different lengths, and section 5 presents bijections between different values of p for fixed words of the same length.

2 Preliminaries

For a word $v \in \{0,1\}^n$, we often write v in the shorthand $v = v_1 v_2 \cdots v_n$ for $v_i \in \{0,1\}$. For a letter $a \in \{0,1\}$, $\overline{a}$ denotes the bitflip of a, $\overline{a} = 1 - a$, and $||v||_a$ denotes the number of digits of v that are equal to a. For a word v, $\overline{v}$ denotes the bitflip of v, or the word $\overline{v} = \overline{v_1 v_2} \cdots \overline{v_n}$. For a pair of binary words x, y, we often write xy for the concatenation of x and y. For $a \in \{0,1\}$, we use the shorthand a^n for the word consisting of n copies of a.

Definition 2. *For binary words v, w, let $\Omega_{v,w}(v)$ be the set of all binary words $u = xv$ such that v appears exactly once as a subword of u (as the suffix). Define $\Omega_{w,v}(w)$ similarly, and set $\Omega_{v,w} = \Omega_{v,w}(v) \cup \Omega_{v,w}(w)$. Also, set $\Omega^*_{v,w}(v) = \{u : uv \in \Omega_{v,w}(v)\}$, where uv is the concatenation of words u and v, similarly for $\Omega^*_{v,w}(w)$, and $\Omega^*_{v,w} = \Omega^*_{v,w}(v) \cup \Omega^*_{v,w}(w)$.*

$(\Omega_{v,w}, \mathbb{P}_p)$ is the natural probability space on which the random variables τ_v and τ_w are defined in the context of Penney's ante.

Definition 3. *(Overlap sets and polynomials) Given words v and w, the overlap set $\mathcal{O}(v,w)$ is the set of prefixes of word w that are also suffixes of word v:*

$$\mathcal{O}(v,w) = \{w_1 w_2 \cdots w_r : v_{m-r+1} v_{m-r+2} \cdots v_m = w_1 w_2 \cdots w_r\}. \tag{2.1}$$

For any $p \in (0,1)$, the correlation polynomial vw_p is the real number

$$vw_p = \sum_{u \in \mathcal{O}(v,w)} \mathbb{P}_p(u)^{-1} = \sum_{u \in \mathcal{O}(v,w)} p^{-||u||_1} (1-p)^{-||u||_0}. \tag{2.2}$$

Simply put, $\mathcal{O}(v,w)$ is the set of words u that are both a suffix of v and a prefix of w. The overlap polynomials pop out of a clever martingale construction, sometimes called the 'Abracadabra' martingale(s) [7]. By a slight abuse of notation, we use xy for concatenation of words, and xy_p for the correlation polynomial value, since we rely on the former heavily, and the latter is common in the literature.

Our main tool is the following formula, originally derived by Conway in unpublished work.

Theorem 4 (Win probability)**.** *Let v, w be any words over the binary alphabet $\{0,1\}$. Then*

$$\mathbb{E}_p \tau_v = vv_p, \tag{2.3}$$

and if neither v nor w is a subword of the other,

$$\mathrm{Win}(v, w; p) := \mathbb{P}_p(\tau_v < \tau_w) = \mathbb{P}_p(\Omega_{v,w}(v)) = \frac{ww_p - wv_p}{ww_p + vv_p - wv_p - vw_p}. \tag{2.4}$$

It is not possible for arbitrary sets to occur as overlap sets. We will use the following special property of overlap sets, which is a simple exercise:

Theorem 5 (Forward propagation [5]). *Let v be a binary word of length n. The set*

$$\{n - k : v_1 v_2 \cdots v_k \in \mathcal{O}(v, v)\} \tag{2.5}$$

is closed under addition.

Additionally, we will use the following 'bad prefix' sets:

Definition 6. *For words v, w, let $\mathcal{B}(v, w)$ denote the set of prefixes of v corresponding to non-trivial overlaps of v, w: that is, $\mathcal{B}(v, w)$ consists of all words $v_1 \cdots v_{n-k}$ such that $w_1 \cdots w_k = v_{n-k+1} \cdots v_n$, i.e. $w_1 \cdots w_k \in \mathcal{O}(v, w)$. Also, let $\mathcal{F}(v, w) := \mathcal{B}(v, v) \cup \mathcal{B}(w, w) \cup \mathcal{B}(w, v) \cup \mathcal{B}(v, w)$.*

Note that $\mathcal{O}(v, w)$ and $\mathcal{B}(v, w)$ are not symmetric in (v, w), but $\mathcal{F}(v, w)$ is. To clarify the definitions, we work out everything for an explicit example.

Example 7.

3 Overview & Results

Our first main result, proved in Section 4, characterizes when a longer word can beat a shorter one for $p = \frac{1}{2}$.

Theorem 8. *Let v and w be words over the alphabet $\{0, 1\}$ with $len(v) = len(w) + 1$, and such that v does not contain w as a subword. If $Win\left(v, w; \frac{1}{2}\right) \geq \frac{1}{2}$, then $w \in \{1^n, 0^n\}$.*

Since the words 1^n and 0^n have the largest possible expected hitting time of all words of length n, it is intuitive that longer words can have the best chance against these two candidates. A common heuristic appearing in previous work is that to optimize the win probability against a fixed word w, you should take a long prefix of w and make it a suffix of v [4]. Heuristically, this is a good choice because sometimes, when w is about to occur, v has just occurred.

Our other main results, detailed in Section 5, describe special properties of the win probability. For a fixed pair v, w of words with the same length, we first show that the limit $p \to 0$ can only give three possible win probabilities:

Proposition 9. *For any words v, w, $\lim_{p \to 0} \mathrm{Win}(v, w; p) \in \{0, \frac{1}{2}, 1\}$. Moreover, the limit is $\frac{1}{2}$ if and only if $||v||_1 = ||w||_1 = r$ and neither $\mathcal{O}(v, w)$ nor $\mathcal{O}(w, v)$ contains an overlap word u such that $||u||_1 = r$.*

The condition in the second part of Proposition 9 should be thought of as a weaker form of the condition that neither v nor w contains the other as a subword. In a nutshell, it says that for $p \approx 0$, v and w do not occur almost at the same time with high probability.

Next, we focus on three possible symmetries of the function $\mathrm{Win}(v, w; p)$. Say a pair v, w has *even*, *odd*, or *constant symmetry* if $\mathrm{Win}(v, w; \frac{1}{2} + x)$ is an even, odd, or constant function for $x \in (-\frac{1}{2}, \frac{1}{2})$, respectively. Note that by Proposition 9, constant symmetry can only occur if

$\mathrm{Win}(v, w; p) = \frac{1}{2}$ for all $p \in (0, 1)$ (if $v \neq w$). While these symmetries can occur for trivial reasons, for example when $w = \overline{v}$, most of the time it is not obvious why we observe such a symmetry. Our main results, Theorems 10 and 12, identify large classes of pairs where these symmetries occur, and give explicit bijections which explain the symmetry.

Theorem 10. *Let $\mathcal{R}$ denote the family of pairs of binary words (v, w) of the same length such that $\Omega^*_{v,w}(v) = \Omega^*_{v,w}(w)$.*

i. For $(v, w) \in \mathcal{R}$,

$$\mathrm{Win}(v, w; p) = \frac{\mathbb{P}_p(v)}{\mathbb{P}_p(v) + \mathbb{P}_p(w)} = \frac{p^s}{p^s + (1-p)^s}, \tag{3.1}$$

where $s = ||v||_1 - ||w||_1$.

ii. If $(v, w) \in \mathcal{R}$, then (v, w) has odd symmetry; and if in addition $||v||_1 = ||w||_1$, then (v, w) has constant symmetry.

iii. The condition $(v, w) \in \mathcal{R}$ can be checked in polynomial time (with respect to the length of the words).

The condition $\Omega^*_{v,w}(v) = \Omega^*_{v,w}(w)$ says that the copy of v (resp. w) that occurs right at the end of the string race can be replaced with w (resp. v) without creating additional copies of v or w. Thus in this case the bijection we seek is simply the identity map on $\Omega^*_{v,w}(v) = \Omega^*_{v,w}(w)$. A computer check suggests that property $\mathcal{R}$ is macroscopically common among all pairs of words:

Conjecture 11. *Let $\mathcal{R}_n$ denote the sub-family of pairs in $\mathcal{R}$ of length n. The limit*

$$\lim_{n\to\infty} 4^{-n} |\mathcal{R}_n| \tag{3.2}$$

exists and is strictly positive.

It is relatively easy to see that if v and w have trivial self and cross overlap sets, i.e. if $\mathcal{O}(v, w) = \mathcal{O}(w, v) = \emptyset$ and $\mathcal{O}(w, w) = \mathcal{O}(v, v) = \{n\}$, then by a direct computation, the Conway formula gives the conclusions i and ii (and $(v, w) \in \mathcal{R}$). However, an exact computer search with $n = 15$ returned the approximate value .15753 for the limit in Conjecture 11, while the proportion of pairs that have trivial auto- and cross- overlap sets is approximately .02309. Thus, Theorem 10 gives a much stronger result than than one obtains by a naive application of the Conway formula. The pair $v = 000100, w = 001110$ is an example where property $\mathcal{R}$ holds, but the auto- and cross-correlations are non-trivial. So there is some non-obvious cancellation going on in the Conway formula to give us the simple quotient in Theorem 10.

Still, property $\mathcal{R}$ is not necessary to have odd or constant symmetry. The minimal example is 1000 and 0110: this pair has odd symmetry, with $\mathrm{Win}(1000, 0110; p) = p$ – which is what one would obtain from Equation 3.1 – but $(v, w) \notin \mathcal{R}$. Oddly, $(v, \overline{w}) \in \mathcal{R}$, but this seems to be a coincidence. We do not know of a simple bijection that explains the odd symmetry in this case. See Section 6 for further discussion.

The minimal example $(v, w) \notin \mathcal{R}$ with constant symmetry is $(01100101, 01010110)$, which we discovered in a computer search. We explain the symmetry for this example by an explicit, new bijection in Section 5, and generalize that construction to obtain a large family of such pairs:

Theorem 12. *There exists a family $\mathcal{E}$ of pairs of words (v, w) of the same length such that the following hold:*

i. *For* $(v,w) \in \mathcal{E}$, *there exists a bijection* $\varphi : \Omega_{v,w}(v) \to \Omega_{v,w}(w)$ *such that* $\mathbb{P}_p(x) = \mathbb{P}_p(\varphi(x))$. *In particular,* (v,w) *has constant symmetry.*

ii. *The condition* $(v,w) \in \mathcal{E}$ *can be checked in polynomial time (with respect to the length of the words).*

iii. $\mathcal{E} \setminus \mathcal{R}$ *is infinite.*

Remark: We note (but do not prove) that the map φ has a *finite expected coding radius,* meaning: given a random $X \in \Omega_{v,w}(v)$ distributed according to the measure $\mathbb{P}_p$ conditioned on the event $\tau_v < \tau_w$, any digit of the image word $\varphi(x)_i$ can be determined by examining a block $x_j x_{j+1} \cdots X_{j+R-1}$ of random size R with $\mathbb{E}R < \infty$ (where the expectation is with respect to the conditioned measure). In the language of symbolic dynamics, this is the next best thing after being a conjugacy-like map, i.e. if R were finite almost surely.

Finally, we turn to even symmetry, which we understand the least. By using a graphical representation of Penney's ante, we construct an explicit bijection for the pair $(1010, 1100)$, which has even symmetry. We do not know how to generalize this construction to a larger family of pairs – see Section 6 for further discussion.

4 Long vs short

Our starting point is the proof of Theorem 8, which characterizes when longer words can be favorable over shorter ones in the unbiased case $p = \frac{1}{2}$.

Proof of Theorem 8. We prove the contrapositive statement, so assume $w \neq 1^n$ or 0^n. For simplicity, we write $xy = xy_{\frac{1}{2}}$ for the auto- or cross correlation polynomial, with $x, y \in \{v, w\}$. By Theorem 4 and a bit of simple algebra,

$$\mathrm{Win}(v,w;p) < \frac{1}{2} \iff ww - wv < vv - vw. \tag{4.1}$$

Note that $n-1 \notin \mathcal{O}(w,w)$, since otherwise by Theorem 5, $\mathcal{O}(w,w) = \{1, 2, \ldots, n-1, n\}$, which is easily seen to be equivalent to $w = 1^n$ or 0^n.

We now divide into a few simple cases, and establish equation 4.1 in each one.

Case 1 Suppose $n-1 \notin \mathcal{O}(v,w)$. Then

$$vw \leq \sum_{1 \leq i \leq n-2} 2^i = \frac{1}{2} 2^n. \tag{4.2}$$

If $n-2 \in \mathcal{O}(w,w)$, then $n-x \notin \mathcal{O}(w,w)$ for all odd integers x, since otherwise by Theorem 5, we would have $w_1 = w_3 = \cdots$, $w_2 = w_4 = \cdots$, and $w_1 = w_{x+1}$, which altogether imply $w = 1^n$ or 0^n. Thus we have

$$ww \leq \left(\sum_{\substack{1 \leq i \leq n \\ i \text{ even}}} 2^i \right) \wedge \left(\sum_{\substack{1 \leq i \leq n \\ i \neq n-2, n-1}} 2^i \right) \leq \frac{4}{3} 2^n. \tag{4.3}$$

We always have the bounds $vv \geq 2^{n+1}$ and $wv \geq 0$, so in this case we obtain

$$ww - wv \leq \frac{4}{3}2^n \text{ and } vv - vw \geq 2^{n+1} - \frac{1}{2}2^n = \frac{3}{2}2^n. \tag{4.4}$$

Since $\frac{3}{2} > \frac{4}{3}$, this gives the desired bound.

Case 2 Suppose $n-1, n-2 \in \mathcal{O}(v,w)$. Unraveling the definition of $\mathcal{O}(v,w)$ gives

$$v = v_1v_2w_1w_2\cdots w_{n-1} = v_1v_2v_3w_1w_2\cdots w_{n-2} \text{ and thus } w = w_1w_1\cdots w_1w_n. \tag{4.5}$$

Since $w \neq 1^n$ or 0^n, $w_n \neq w_1$, so $ww = 2^n$. Since v does not contain w as a subword, $n \notin \mathcal{O}(v,w)$, so $vw \leq \sum_{i=1}^{n-1} 2^i \leq 2^n$. Therefore

$$ww - wv \leq 2^n - 0 = 2^n, \text{ and } vv - vw \geq 2^{n+1} - 2^n = 2^n, \tag{4.6}$$

as desired.

Case 3 Suppose $n-1 \in \mathcal{O}(v,w)$, $n-2 \notin \mathcal{O}(v,w)$, and $n-2 \notin \mathcal{O}(w,w)$. Then we bound ww and vw directly:

$$vw \leq \sum_{\substack{1\leq i\leq n-1 \\ i\neq n-2}} 2^i \leq \frac{3}{4}2^n, \tag{4.7}$$

and

$$ww \leq \sum_{\substack{1\leq i\leq n \\ i\neq n-2,n-1}} 2^i \leq 2^n + 2^{n-2} = \frac{5}{4}2^n. \tag{4.8}$$

Combining these as before, with $vv \geq 2^{n+1}$ and $wv \geq 0$ gives the desired bound.

Case 4 Suppose $n-1 \in \mathcal{O}(v,w)$, $n-2 \notin \mathcal{O}(v,w)$, and $n-2 \in \mathcal{O}(w,w)$. Then $w_1 = w_3 = \cdots$ and $w_2 = w_4 = \cdots$, so without loss of generality,

$$w = \begin{cases} (10)^{\frac{n}{2}}, & n \text{ even} \\ (10)^{\frac{n-1}{2}}1, & n \text{ odd} \end{cases} \tag{4.9}$$

and

$$v = \begin{cases} v_1v_2(10)^{\frac{n}{2}-1}1, & n \text{ even} \\ v_1v_2(10)^{\frac{n-1}{2}}, & n \text{ odd} \end{cases} \tag{4.10}$$

In both cases, by a direct computation,

$$vw = \sum_{\substack{1\leq i\leq n-2 \\ i \text{ odd}}} 2^{n-i} \leq \frac{2}{3}2^n \tag{4.11}$$

and

$$ww = \sum_{\substack{0 \leq i \leq n-1 \\ i \text{ even}}} 2^{n-i} \leq \frac{4}{3} 2^n. \tag{4.12}$$

Combining these as before, with $vv \geq 2^{n+1}$ and $wv \geq 0$ gives the desired bound.

□

Theorem 8 is tight, in the sense that the inequality Win $\geq \frac{1}{2}$ cannot be replaced by Win $\geq \frac{1}{2} - \epsilon$ for any $\epsilon > 0$, since (for example) the pair $v = 11(10)^m$, $w = (10)^m 1$ achieves $q_{w,v}\left(\frac{1}{2}\right) = \frac{1}{2} - O(2^{-2m})$. (This is the same pair that appears in **Case 4** of the proof of Theorem 8.)

For any words with length difference $k \geq 1$, the optimal win probability (for the longer word) always occurs when the shorter word is 1^n or 0^n:

Theorem 13. *Let v and w be such that len(v) = len$(w) + k$ for non-negative integer k, and v does not contain w as a subword. Then*

$$\text{Win}\left(v, w; \frac{1}{2}\right) < \frac{2}{1 + 2^k} \tag{4.13}$$

and equality achieved asymptotically for the pair $v = 0^{k+1}1^{n-1}, w = 1^n$.

Note the $k = 0$ case is already a result of Felix [1] J: the $k = 0$ case is really easy, so I guess he proves more there?. To prove proposition 13, we just need to generalize the bound on q that we did in Theorem 8 to other values of k.

Proof. We only need crude bounds on each of the overlap polynomials vv, ww, wv, vw: namely, $vv \geq 2^{n+k}$, $ww \leq \sum_{i=1}^{n} 2^i \leq 2^{n+1}$, $wv \geq 0$, and since v is not a subword of w, $n \notin \mathcal{O}(v, w)$, so $vw \leq \sum_{i=1}^{n-1} 2^i \leq 2^n$. Thus, by Theorem 4,

$$\text{Win}\left(v, w; \frac{1}{2}\right) \leq \frac{2^{n+1}}{2^{n+k} + 2^{n+1} - 2^n} = \frac{2}{1 + 2^k}. \tag{4.14}$$

The pair $v = 0^{k+1}1^{n-1}$, $w = 1^n$ asymptotically achieves each of the bounds on vv, ww, wv and vw up to negligible $o(1)$ terms (should this be $O(1)$? Not sure if notation is different in math vs. computer science).

□

4.1 Long vs short with bias

The case $p \neq \frac{1}{2}$ is qualitatively different from $p = \frac{1}{2}$ in the following way. At $p = \frac{1}{2}$, all words have the same probability under $\mathbb{P}_p$, while they don't when $p \neq \frac{1}{2}$. Thus, words with higher probability under $\mathbb{P}_p$ will have an advantage in Penney's ante, even if they are much longer than their opponent. But if two words have comparable probabilities under $\mathbb{P}_p$, then the overlap patterns will be the key feature that determines the winner. This suggests restricting our attention to pairs (v, w) whose probabilities are comparable under $\mathbb{P}_p$ for all p. One natural way to achieve this is to require v and w to have the same number of 1s, and take $p \in (0, \frac{1}{2})$. Thus we seek to optimize over the following set:

Definition 14. *For any $n \geq 2$ and $k \geq 1$, let $W(n, k)$ denote the set of pairs (v, w) of words of length $n + k$ and k respectively such that $||v||_1 = ||w||_1$ and w is not a subword of v.*

Computer calculations suggest that the following version of Theorem 13 holds.

Conjecture 15. *The maximum value of $Win(v, w; p)$ over the set $W(n, k)$ can be computed as follows. There are two cases, depending on the value of k:*

$(k = 1)$ *There exists $r = r(n) \in (0, \frac{1}{2})$ such that if $p \le r(n)$, the maximum occurs for the pair $(001^{n-1}, 1^{n-1}0)$; and if $p \ge r(n)$, the maximum occurs for the pair $(101^{n-1}, 1^n)$.*

$(k \ge 2)$ *There exists $r' = r'(n, k) \in (0, \frac{1}{2})$ such that if $p \le r'$, the maximum occurs for pair $(0^{k-1}101^{n-1}, 1^n)$; and if $p \ge r'$, the maximum occurs for the pair $(0^{k+1}(10)^{m-1}1, (10)^m)$ where $n = 2m$ is even, and for the pair $(0^{k+2}(10)^{m-1}1, 0(10)^m)$ if $n = 2m + 1$ is odd.*

We verified conjecture 15 by computer computations for all pairs of words with $n \le 11$ and $k \le 4$. It is surprising that only words that are periodic (or nearly periodic) with period 1 or 2 appear as the optimizers. A back of the envelope calculation suggests that pairs where w has a longer period do not give higher win probabilities, but we do not know a proof. Assuming the conjecture, we can precisely compute the maximum win probability for longer words, and the transition point between (pseudo-)period 1 and 2 optimizers:

Corollary 16. *Assume Conjecture 15 holds, and recall the notation there. For $k = 1$, $p \in (0, \frac{1}{2})$ and $v, w \in W(n, k)$,*

$$Win(v, w; p) \le \begin{cases} (1-p)(1-p^{n-1}), & p \in (0, r] \\ \frac{1-p^{n-1}}{2-p}, & p \in [r, \frac{1}{2}) \end{cases} \tag{4.15}$$

The threshold $r = r(n)$ satisfies

$$r(n) = \kappa + O(\kappa^n) \text{ as } n \to \infty, \tag{4.16}$$

where $\kappa = \frac{3-\sqrt{5}}{2}$. For $n, k \ge 2$, any $p \in (0, \frac{1}{2})$ and $(v, w) \in W(n, k)$,

$$Win(v, w; p) \le \begin{cases} \frac{(1-p)^{k-1}}{1+(1-p)^k}, & p \in [0, r'] \\ \frac{(1-p)^k}{1-p(1-(1-p)^k)+p^2}, & p \in [r', \frac{1}{2}) \end{cases} \tag{4.17}$$

For any $k \ge 2$, the threshold $r' = r'(n, k)$ satisfies

$$\lim_{n\to\infty} r'(n, k) = \sqrt{2} - 1. \tag{4.18}$$

We omit the proofs, which are straightforward calculations applying Theorem 4 to the explicit pairs appearing in Conjecture 15. Note that the optimal win probability only depends on n when $k = 1$, and only on k if $k \ge 2$ and n is even. For the case where $k \ge 2$ and n is odd, the optimal pair $(0^{k+2}(10)^{m-1}1, 0(10)^m)$ has win probability which strictly increases as $n \to \infty$ to the value for the n even case.

5 Symmetry and bijections

In this section we investigate symmetries that underlie the combinatorics of Penney's ante. We restrict our attention to words v and w with the same length. The most general question we could seek to answer is: for two values $p, p' \in (0, 1)$ and for two pairs $(v, w), (v', w')$, when is there a coupling between the v vs w race under $\mathbb{P}_p$ and the (v', w') race under $\mathbb{P}_{p'}$? Here we restrict our attention to comparing a fixed pair (v,w) to itself, but under different values of p. Our primary aim is to characterize when symmetries of the function $\mathrm{Win}(v,w;p)$ occur, and to construct explicit bijections that explain the symmetry. Recall the terminology from Section 3: the pair (v,w) has *odd symmetry* if $\mathrm{Win}(v,w;1-p)=1-\mathrm{Win}(v,w;p)$ for $p\in(0,1)$; *even symmetry*, if $\mathrm{Win}(v,w;1-p)=\mathrm{Win}(v,w;p)$ for $p\in(0,1)$; and *constant symmetry* if $\mathrm{Win}(v,w;p)=\frac{1}{2}$ for all $p\in(0,1)$.

As a warm-up, we prove Proposition 9.

Proof of Proposition 9. First, observe that a term of the form $p^{-||w||_1}(1-p)^{-j}$ for non-negative integer j occurs exactly once in the overlap polynomial ww_p, when $j=||w||_0$. Similarly, there is at most one term of the form $p^{-||v||_1}(1-p)^{-k}$ in vv_p. Also, at most one of vw_p and wv_p may contain a term of that form: if there exists $u\in\mathcal{O}(v,w)$ with $||u||_1=r$, then we must have $v=0^a1u'0^b$ and $w=0^c1u'0^d$ for some non-negative a,b,c,d with $c\le a$, and where u' is a subword of u. By the assumption that neither v nor w contains the other as a subword, $d>b$, which implies that $\mathcal{O}(w,v)$ does not have an overlap with r 1s.

Using formula 4, the win probability is a rational function of the form

$$\mathrm{Win}(v,w;p)=\frac{\sum_{0\le i,j\le m}(a_{ij}-b_{ij})p^{-i}(1-p)^{-j}}{\sum_{0\le i,j\le m}(a_{ij}-b_{ij}+c_{ij}-d_{ij})p^{-i}(1-p)^{-j}},\tag{5.1}$$

where the a,b,c,d coefficients are $\{0,1\}$ valued. Let r denote the maximum power of p^{-1} that appears in either sum. By the observations in the previous paragraph, $r=||w||_1\vee||v||_1$, the coefficient of p^{-r} in the numerator is 0 or 1, and the coefficient of p^{-r} in the denominator is either 1 or 2 (up to some bounded number of factors of $1-p$, which are $1+O(p)$ as $p\to 0$). Thus the only possible $p\to 0$ limits are the quotients $\{\frac{0}{1},\frac{0}{2},\frac{1}{1},\frac{1}{2}\}=\{0,\frac{1}{2},1\}$.

For the second part, if $||v||_1\ne||w||_1$, then the denominator has leading coefficient 1 and the numerator has leading coefficient either 0 or 1, so the limit must be 0 or 1. Similarly, if $||v||_1=||w||_1=r$ and one of $\mathcal{O}(v,w)$ or $\mathcal{O}(w,v)$ has an overlap that contains r 1s, then the leading coefficient in the numerator is 0 or 1, and the leading coefficient in the denominator is 1, so the limit cannot be $\frac{1}{2}$.

□

5.1 Odd Symmetry

Recall we say odd symmetry holds if $\mathrm{Win}(v,w;p)=1-\mathrm{Win}(v,w;1-p)$ for all $p\in(0,1)$, i.e. if

$$\mathbb{P}_p(\tau_v<\tau_w)=1-\mathbb{P}_{1-p}(\tau_v<\tau_w)=\mathbb{P}_{1-p}(\tau_w<\tau_v)=\mathbb{P}_p(\tau_{\overline{w}}<\tau_{\overline{v}}).\tag{5.2}$$

Our aim is to construct canonical bijections that give rise to this equality. A class of pairs that obviously has odd symmetry is when $v=\overline{w}$. This odd symmetry is naturally realized via the bijection $\varphi:\Omega_{v,\overline{v}}(v)\to\Omega_{v,\overline{v}}(\overline{v})$ given by $\varphi(u)=\overline{u}$. Since $\mathbb{P}_p(\varphi(u))=\mathbb{P}_{1-p}(u)$, this bijection induces the equality $\mathbb{P}_p(\tau_v<\tau_{\overline{v}})=\mathbb{P}_{1-p}(\tau_{\overline{v}}=\tau_v)$. Next we turn to a much larger class of pairs with odd symmetry.

Definition 17 (Property $\mathcal{R}$)**.** *We say a pair of binary words (v,w) has property $\mathcal{R}$, and write $(v,w)\in\mathcal{R}$, if $\Omega^*_{v,w}(v)=\Omega^*_{v,w}(w)$.*

Simply put, property $\mathcal{R}$ says that at the end of the Penney's ante race, swapping the final v to a w or vice versa does not create any additional copies of v or w. To check whether property $\mathcal{R}$ holds, it is not necessary to directly check if the infinite sets $\Omega^*_{v,w}(v)$ and $\Omega^*_{v,w}(w)$ are equal. Instead, one only needs to examine the 'bad prefix' set $\mathcal{F}$:

Lemma 18. *A pair of words (v, w) has property $\mathcal{R}$ if and only if $fv, fw \notin \Omega_{v,w}$ for all $f \in \mathcal{F}(v, w)$.*

Proof. Let n be the length of v, w. We will justify the following chain of equivalences:

$$\begin{aligned}
\mathcal{O}^*_{v,w}(v) = \mathcal{O}^*_{v,w}(w) &\iff \text{for all } x \in \Omega^*_{v,w}, (xv \notin \Omega_{v,w} \iff xw \notin \Omega_{v,w}) &(5.3)\\
&\iff \text{for all } x \in \{0,1\}^{n-1}, (xv \notin \Omega_{v,w} \iff xw \notin \Omega_{v,w}) &(5.4)\\
&\iff \text{for all } f \in \mathcal{F}(v,w), (fv \notin \Omega_{v,w} \iff fw \notin \Omega_{v,w}) &(5.5)\\
&\iff \text{for all } f \in \mathcal{F}(v,w), (fv, fw \notin \Omega_{v,w}) &(5.6)
\end{aligned}$$

The equivalence 5.3 is clear. For 5.4, the $\implies$ implication is clear since the latter condition is weaker than the former. For the reverse implication, we used the fact that since $x \in \Omega^*_{v,w}$ does not have v or w as a subword, no v or w can occur in xv or xw without intersecting the final $n-1$ digits of x, along with the fact that $\Omega^*_{v,w}$ contains all words of length $n-1$. For 5.5, the $\implies$ implication is again clear since the latter condition is weaker than the former. For the reverse implication, note that if for example $xv \notin \Omega_{v,w}$, then xv contains either a copy of w or two copies of v, which implies that some suffix f of x is an element of $\mathcal{D}(v,v) \cup \mathcal{D}(w,v)$, and similarly for xw. The final equivalence follows from the fact that $fv, fw \in \Omega_{v,w}$ for any $f \in \{0,1\}^{n-1} \setminus \mathcal{F}(v,w)$. □

Proposition 19. *Suppose $v \neq w$ are any words of the same length such that (v, w) satisfies property $\mathcal{R}$. Then for all $p \in (0, 1)$,*

$$Win(v, w; p) = \frac{\mathbb{P}_p(v)}{\mathbb{P}_p(v) + \mathbb{P}_p(w)}. \tag{5.7}$$

In particular, the pair (v, w) is odd symmetric.

In words: if v and w have this 'valid replacement' property, then their relative odds of one winning in Penney's ante is just their probabilities as sequences under the iid measure $\mathbb{P}_p$.

Proof. By property $\mathcal{R}$, let $\Omega^* = \Omega^*_{v,w}(v) = \Omega^*_{v,w}(w)$. By the definition of $\Omega^*_{v,w}(v)$ and using that $\mathbb{P}_p$ is iid,

$$\text{Win}(v, w; p) = \mathbb{P}_p(\Omega_{v,w}(v)) = \mathbb{P}_p(\Omega^*)\mathbb{P}_p(v), \tag{5.8}$$

and similarly,

$$\mathbb{P}_p(\Omega_{v,w}(w)) = \mathbb{P}_p(\Omega^*)\mathbb{P}_p(w). \tag{5.9}$$

Since $v \neq w$, $\mathbb{P}_p(\Omega_{v,w}(v)) + \mathbb{P}_p(\Omega_{v,w}(w)) = 1$, and thus

$$\text{Win}(v, w; p) = \frac{\mathbb{P}_p(\Omega^*)\mathbb{P}_p(v)}{\mathbb{P}_p(\Omega^*)\mathbb{P}_p(v) + \mathbb{P}_p(\Omega^*)\mathbb{P}_p(w)} = \frac{\mathbb{P}_p(v)}{\mathbb{P}_p(v) + \mathbb{P}_p(w)}. \tag{5.10}$$

Straightforward algebra shows that this implies odd symmetry.

□

5.2 Constant symmetry

Recall that we say a pair v, w has constant symmetry if $\text{Win}(v, w; p) = \frac{1}{2}$ for all $p \in (0, 1)$. Note that Proposition 10 guarantees a large class of pairs with constant symmetry: (v, w) has constant symmetry whenever $||v||_1 = ||w||_1$ and property $\mathcal{R}$ holds for the pair (v, w). (An example is $0101, 0110$.) However, property $\mathcal{R}$ is not necessary for constant symmetry. We start by detailing the minimal example, which will easily generalize to an infinite class of such examples.

Lemma 20. *Let $a = 0110, b = 0101$, and $v = 01100101 = ab, w = 01010110 = ba$. There exists a bijection $\varphi : \Omega_{v,w}(v) \to \Omega_{v,w}(w)$ such that for any $p \in (0, 1)$, $\mathbb{P}_p(x) = \mathbb{P}_p(\varphi(x))$. The pair (v, w) has constant symmetry.*

Note that property $\mathcal{R}$ does not hold for v, w, since $a \in \Omega^*_{v,w}(v)$, but $a \notin \Omega^*_{v,w}(w)$. The rough idea is: we want to construct a bijection that swaps occurrances of a and b. Although there are many candidate maps that seem to achieve this, we only found one bijection that works.

Proof. Note that each word $z \in \Omega_{v,w}(v)$ can be uniquely expressed in the form $z = xa^jv \in \Omega_{v,w}(v)$ where $j \geq 0$ is maximal. Write $j = j(z)$ for the maximal such j. Then define the map $\varphi_{a,b}$ by

$$\varphi_{a,b}(xa^jv) = xb^jw, \tag{5.11}$$

and define $\varphi_{b,a}$ similarly, with the roles of a and b reversed. If $\varphi_{a,b}$ is indeed a well-defined map from $\Omega_{v,w}(v)$ to $\Omega_{v,w}(w)$, and similarly in the reverse for $\varphi_{b,a}$, then $\varphi_{a,b}$ is the desired map: indeed, we have $\mathbb{P}_p(a) = \mathbb{P}_p(b)$, and thus $\mathbb{P}_p(v) = \mathbb{P}_p(w)$, so both $\varphi_{a,b}$ and $\varphi_{b,a}$ are clearly measure-preserving, and $\varphi_{a,b} = \varphi_{b,a}^{-1}$. We will show that $\varphi_{a,b}$ is well-defined; the details for $\varphi_{b,a}$ are essentially the same.

Write $\Omega_{v,w} = \Omega$ for short, and for a word u, let $u_{[i]} = u_1u_2 \cdots u_i$ denote the i prefix of u. Our goal is to show that if $yv \in \Omega(v)$, then $\varphi(yv) \in \Omega(w)$. The key observation is that for any word y that does not contain v or w as a subword,

$$yv \in \Omega(v) \iff y \neq y'u \text{ for } u \in \mathcal{D}(v, v) \cup \mathcal{D}(w, v) = \{b, v_{[6]}, w_{[7]}\} \tag{5.12}$$

and

$$yw \in \Omega(w) \iff y \neq y'u \text{ for } u \in \mathcal{D}(w, w) \cup \mathcal{D}(v, w) = \{a, v_{[6]}, w_{[7]}\}. \tag{5.13}$$

Consider a word $z = xa^jv \in \Omega(v)$ with $j(z) = 0$, i.e. a word $xv \in \Omega(v)$ where x does not end in a. By statement 5.12, x does not end in $a, b, v_{[6]}$ or $w_{[7]}$; and then by statement 5.13, $xw \in \Omega(w)$ and x does not end in b. So $\varphi_{a,b}$ is well-defined on the set of $xv \in \Omega(v)$ with $j(xv) = 0$.

Now consider a word $z = xa^jv \in \Omega(v)$ with $j(z) \geq 1$. Since j is maximal, by statement 5.12, x does not end in $a, b, v_{[6]}$ or $w_{[7]}$. Observe that

$$\mathcal{D}(v, bb) = \{a, v_{[6]}\} \text{ and } \mathcal{D}(w, bb) = w_{[7]}. \tag{5.14}$$

Since we assumed x doesn't end in any of these words, we obtain that $xb^jw = xb^{j+1}a$ doesn't contain any copy of v or w that includes digits of x. Finally, a direct check of the words b^3 and $bba = bw$ shows that neither v nor w can occur as a subword of b^j or bw. Combining all this shows $xb^jw \in \Omega(w)$, as desired.

Since $\mathbb{P}_p(v) = \mathbb{P}_p(w)$, the existence of φ implies (by the same calculation as in the proof of Proposition 10) that $\text{Win}(v, w; p) = \frac{1}{2}$. $\square$

The construction in Lemma 20 generalizes immediately to the following class of words.

Definition 21 (Property $\mathcal{E}$)**.** *We say a pair of binary words (v, w) has property $\mathcal{E}$, and write $(v, w) \in \mathcal{E}$, if there exist words a, b and a finite set S of words such that the following hold:*

I. $||a||_1 = ||b||_1$

II. $v = ab, w = ba$

III. $D(v, v) \cup D(w, v) = S \cup \{a\}$ *and* $D(w, w) \cup D(v, w) = S \cup \{b\}$

IV. *For all* $m \geq 2$, $D(v, a^m) \cup D(w, a^m) \subset S \cup \{a, b\}$ *and* $D(v, a^m) \cup D(w, b^m) \subset S \cup \{a, b\}$

V. *For all* $m \geq 2$, a^m *and* b^m *do not contain* v *or* w *as a subword*

VI. *For all* $m \geq 1$, $a^m b \in \Omega_{v,w}(v)$, $b^m a \in \Omega_{v,w}(w)$.

Proof. (Proof of Theorem 12) The argument in Lemma 20 adapts immediately to this setting: properties II-VI imply, by the same argument, that the map $\varphi_{a,b} : \Omega_{v,w}(v) \to \Omega_{w,v}(w)$ is a well-defined bijection, and property I guarantees it is measure preserving. To see that the condition $(v, w) \in \mathcal{E}$ can be checked in polynomial time in the lengths of v and w, first note that the existence of a, b satisfying II can be determined by checking all prefixes of v, and if some such pair exists, it will be determined by the same algorithm. Given a and b satisfying condition II, conditions I and III require checking linearly many overlap comparisons, and if III holds, the same algorithm will determine the set S. Given a, b and the set S, note that it suffices to check conditions IV, V and VI for $m < M$, where M is at most a linear function of the lengths of a and b. Finally, an explicit computation shows that for all $k \geq 2$, words $a = 01^k0, b = 01^{k-1}01$ give rise to a pair $(a_k b_k, b_k a_k) \in \mathcal{E} \setminus \mathcal{R}$. $\square$

5.3 Even symmetry

We do not know a general construction akin to the families $\mathcal{R}$ or $\mathcal{E}$ which exhibit even (but not constant) symmetry. We study one example in depth, and give an explicit bijection, which may be of interest in its own right.

Proposition 22. *Let $v = 1100, w = 1010$. There exists a bijection $\varphi : \Omega_{v,w}(v) \to \Omega_{v,w}(v)$ such that $\mathbb{P}_p(\varphi(x)) = \mathbb{P}_{1-p}(x)$ for all $p \in (0, 1)$. In particular, (v, w) has even symmetry.*

We note that the existence of such a bijection does not come for free from the even property, or from the Conway formula. To prove Proposition 22, we rely on a certain graph representation of the string race, which appears in Figure 1. We summarize the connection between Penney's ante and this graph in the following lemma:

Lemma 23. *Let $G_{v,w}$ be the graph appearing in Figure 1. The set of paths in $G_{v,w}$ starting at vertex $(\emptyset, \emptyset)$ and ending at $(v, 100)$ is in bijection with $\Omega_{v,w}(v)$, and the set of paths in $G_{v,w}$ starting at state $(\emptyset, \emptyset)$ and ending at $(10, w)$ is in bijection with $\Omega_{v,w}(w)$. Additionally, simple random walk on $G_{v,w}$ started from vertex $(\emptyset, \emptyset)$ and stopped on hitting $\{(v, 100), (10, w)\}$ has the law of $\tau_{v,w}$.*

Proof. Let $\mathcal{Q}(v)$ denote the set of edge-paths in $G_{v,w}$ starting at $(\emptyset, \emptyset)$ and ending at $(v, 100)$. Write $\gamma = (e_1, e_2, \ldots, e_t)$ for such a path in $\mathcal{Q}(v)$, and let π denote the function from the edge set of $G_{v,w}$ to $\{0, 1\}$ which returns the edge label. We claim that the map $\Pi : \mathcal{Q}(v) \to \Omega_{v,w}(v)$ given by

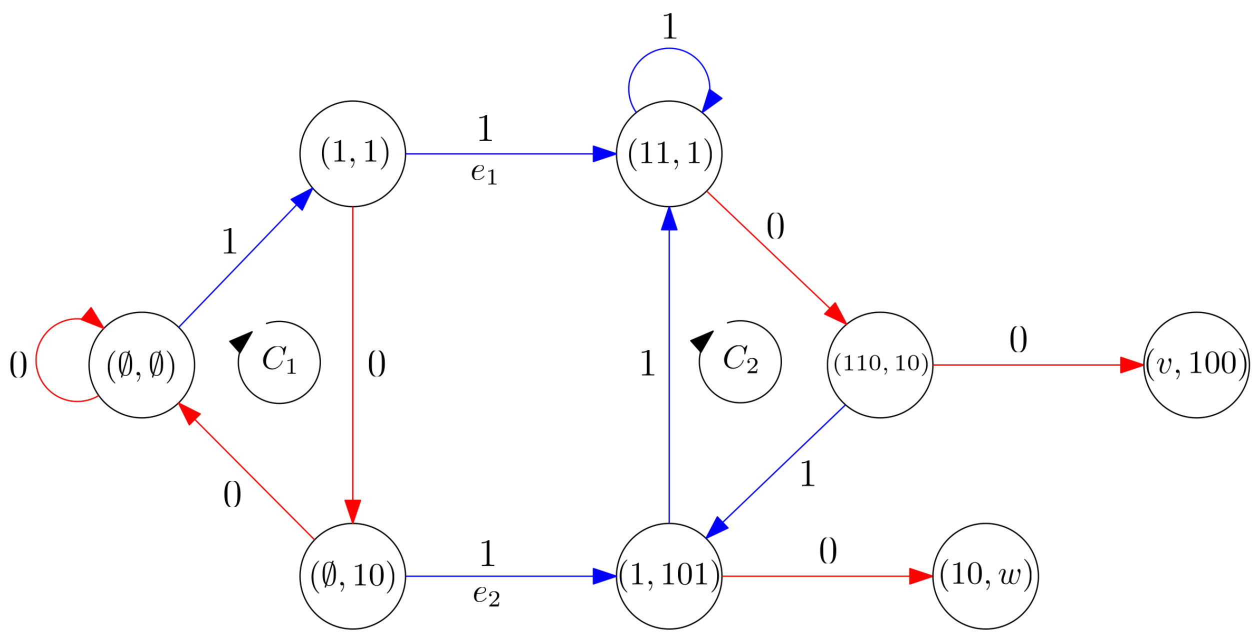

Figure 1: The labeled graph $G_{v,w}$ encoding the space $\Omega_{v,w}$, for $v = 1100, w = 1010$. Elements of $\Omega_{v,w}$ are in bijection with labeled paths in $G_{v,w}$ starting at $(\emptyset, \emptyset)$ and ending at either $(v, 100)$ or $(10, w)$, by reading the labels along the edges of the path. (Blue edges are labeled 0, and red edges are labeled 1.)

Algorithm 1 Construction of $G_{v,w}$

---

queue Stack $\leftarrow [(\emptyset, \emptyset)]$
set Vertices $\leftarrow \{\}$ ▷ Vertices are pairs (v', w') of prefixes of v, w, resp.
set Edges $\leftarrow \{\}$
while len(Stack) > 0 **do**
    $(v', w') \leftarrow$ pop Stack
    for $a \in \{0, 1\}$ **do** ▷ Append a letter on the right to obtain new prefixes
        $v'' \leftarrow \max \mathcal{O}(v'a, v)$ ▷ $\max \mathcal{O}$ denotes the maximum length word in $\mathcal{O}$
        $w'' \leftarrow \max \mathcal{O}(w'a, w)$
        add $(v', w') \rightarrow (v'', w'')$ to Edges ▷ This edge is labelled a
        if $(v'', w'') \notin$ Vertices **then**
            add (v'', w'') to Vertices
            if $v'' \neq v$ **or** $w'' \neq w$ **then**:
                add (v'', w'') to Stack
            end if
        end if
    end for
end while

$$\Pi(\gamma) = \pi(e_1) \cdots \pi(e_t) \tag{5.15}$$

is a bijection. To see why, we explain the construction of $G_{v,w}$ via the algorithmic exploration process in Algorithm 1. In a nutshell, this algorithm generates all words in $\Omega_{v,w}$ by starting from the empty word, appending one letter at a time, and keeping track of the maximal prefixes of v, w that occur as suffixes of the generated word.

Let $z \in \Omega_{v,w}(v)$. By the algorithmic construction, each vertex in $G_{v,w}$ (except the terminal vertices $(v, 100)$ and $(10, w)$) has exactly two outgoing edges, one for each label in $\{0, 1\}$. So there exists a unique edge-path in $G_{v,w}$ obtained by choosing the edges labeled $z_1, z_2, \ldots$ in succession. Since z has v as a suffix, the resulting path ends at vertex $(v, 100)$. Thus Π^{-1} is well-defined, i.e. Π is surjective. By a direct check of all paths of length 4 in the graph $G_{v,w}$, (or alternatively, because of the condition in Line 12 of Algorithm 1), for any $\gamma \in \mathcal{Q}(v)$, $\Pi(\gamma) \in \Omega_{v,w}(v)$. Thus Π is well-defined, and clearly injective, so Π is the desired bijection. It follows from the construction that the law $\mathbb{P}_p$ on $\Omega_{v,w}$ induces the law of simple random walk on $G_{v,w}$.

□

We are now ready to prove Proposition 22.

Proof. Let $\mathcal{P}(v)$ denote the set of vertex-paths $(\emptyset, \emptyset) \to (v, 100)$ in $G_{v,w}$, and for integers $j, k \geq 0$, let $\mathcal{P}_{j,k}(v)$ denote the set of such paths with j edges labeled 0 (red edges) and k edges labeled 1 (blue edges). We will define a bijection $\psi : \mathcal{P}(v) \to \mathcal{P}(v)$, with the property that for all $j, k \geq 0$, ψ restricted to $\mathcal{P}_{j,k}$ is a bijection $\psi_{j,k} : \mathcal{P}_{j,k}(v) \to \mathcal{P}_{k,j}(v)$. Combining this with Lemma 23 gives the result.

For a path $x \in \mathcal{P}_{j,k}(v)$, let $N_1(x)$ (resp. $N_2(x)$) denote the number of times x visits the vertex $(\emptyset, \emptyset)$ (resp. the vertex $(11, 1)$). $N_1 - 1$ and $N_2 - 1$ count the number of times x traverses the cycles C_1 and C_2, up to a small error. Also, for $1 \leq i \leq N_1(x)$, let $s_i^1(x) \in \{0, 1, 2, \ldots\}$ denote the number of times the self loop at $(\emptyset, \emptyset)$ was consecutively traversed by x immediately after the ith visit of x to $(\emptyset, \emptyset)$. Similarly, for $1 \leq i \leq N_2(x)$, let $s_i^2(x)$ denote the number of times the self loop at $(11, 1)$ was consecutively traversed by x immediately after the ith visit of x to $(11, 1)$. Note there is no path from any vertex in the 3-cycle C_2 to any vertex in the 3-cycle C_1, so x traverses exactly one of the two edges e_1, e_2 exactly once. Let $e(x) = 1$ or 2 depending on which edge was traversed. Altogether, any path $x \in \mathcal{P}$ is uniquely described by the tuple

$$(N_1(x), N_2(x), (s_i^1(x))_{1 \leq i \leq N_1(x))}, (s_i^2(x))_{1 \leq i \leq N_2(x))}, e(x)), \tag{5.16}$$

and each such tuple with $N_1, N_2 \geq 1, s_i^1, s_i^2 \geq 0, e \in \{1, 2\}$ determines a unique x. Now we define $\psi(x)$ as the path $y \in \mathcal{P}$ given by the tuple

$$(N_1(y), N_2(y), (s_i^1(y)), (s_i^2(y)), e(y)) = (N_2(x), N_1(x), (s_i^2(x)), (s_i^1(x)), e(x)). \tag{5.17}$$

In words, whatever x does in the cycle C_1, $y = \psi(x)$ does the same in the cycle C_2 and vice versa, and y crosses between the cycles at the same edge that x does. It remains to check that for $x \in \mathcal{P}_{j,k}, y \in \mathcal{P}_{k,j}$. By counting edge labels in $G_{v,w}$ along any path $z \in \mathcal{P}_{r,b}$, one finds

$$r = 2(N_1(z) - 1) + \sum_{i=1}^{N_1(z)} s_i^1(z) + 1\{e(z) = 2\} + N_2(z) + 1, \tag{5.18}$$

and

$$b = N_1(z) + 1 + 1\{e(z) = 2\} + \sum_{i=1}^{N_2(z)} s_i^2(z) + 2(N_2(z) - 1). \tag{5.19}$$

The terms in Equation 5.18 come from: the cycle C_1, which is fully traversed $2(N_1(z) - 1)$ times, and contains two edges labeled 0; the self-loop at $(\emptyset, \emptyset)$, which is traversed $\sum s_i^1(z)$ many times; the edge $(1, 1) \to (\emptyset, 10)$, which is traversed an additional time (not counted by $N_1(z) - 1$) if $e(z) = 2$; the cycle C_2, which is fully traversed $N_2(z)$ times and contains one edge labeled 0; and the final edge $(110, 10) \to v$, which is traversed exactly once. The counting is similar for Equation 5.19.

Observe that Equations 5.18 and 5.19 are exchanged when $N_1(z)$ and $N_2(z)$ are swapped and the sequences $(s_i^1(z))$ and $(s_i^2(z))$ are swapped. Thus, using the definition 5.17 of ψ and equations 5.18 and 5.19 for both $(r, b) = (j, k)$ with path x and $(r, b) = (k, j)$ with path $y = \psi(x)$ completes the proof.

□

6 Discussion

Our work suggests many interesting further directions. While our findings describe properties of the win probability for words of different lengths, and for arbitrary $p \in (0, 1)$, they say little about the optimal strategy in Penney's ante: that is, given p and a word w, which word v optimizes the win probability? A natural next step would be to extend results on the optimal strategy for the $p = 1/2$ case to arbitrary p, or to other underlying markov chains. Proving Conjecture 15 could be a good starting point. Alternatively, it may be better to work with different conditions on pairs: for example, rather than requiring $||v||_1 = ||w||_1$, one could restrict to pairs with $|\mathbb{P}_p(v) - \mathbb{P}_p(w)| < \varepsilon$ or $\frac{\mathbb{P}_p(v)}{\mathbb{P}_p(w)} \in (1 - \varepsilon, 1 + \varepsilon)$ for some ε depending on p and the lengths of v and w. What are the optimal pairings under this type of restriction?

A related question is to extend Theorem 8 in the following way:

Problem 24. *For $\alpha \in (0, 1)$, characterize the set of words $w \in \{0, 1\}^n$ such that $Win(v, w; p) \geq \alpha$ for some $v \in \{0, 1\}^{n+1}$.*

As noted in the introduction, $\mathrm{Win}(v, w; p)$ need not be a monotone function of p for fixed v, w. In general, what kinds of functions $f : [0, 1] \to [0, 1]$ can arise as win probability functions, and which cannot? By Theorem 4 f is rational, and by Theorem 9 $f(0), f(1) \in \{0, \frac{1}{2}, 1\}$. What can be said about the derivative of the win probability as a function of p? Do values of p where $\frac{\partial}{\partial p}\mathrm{Win}(v, w; p) = 0$ give probability spaces with some special or extremal properties? Based on heuristic arguments and computer checks, we believe the following conjecture holds:

Conjecture 25. *There exists $C > 0$ such that for any positive integer k, there exists a pair (v, w) with lengths at most $\exp(Ck)$ such that the derivative $\frac{d}{dp} Win(v, w; p)$ changes sign at least k times for $p \in (0, 1)$.*

Our results for odd symmetry deal with a majority of pairs, but there are still some frustrating outliers. The pair $v = 1000, w = 0110$ is an example with odd symmetry, but where property $\mathcal{R}$ does not hold. It is easy to see that there is no length-preserving bijection $f : \Omega_{v,w}(v) \to \Omega_{\overline{v},\overline{w}}(\overline{w})$ that preserves $\mathbb{P}_p$: indeed, any length-preserving f must have $f(v) = \overline{w}$ (since these are the only allowable words of length 4), but since v and $\overline{w}$ don't have the same number of 1s, f doesn't preserve $\mathbb{P}_p$. The next best thing would be to find a pair of partitions $\mathcal{A}$ of $\Omega_{v,w}(v)$ and $\mathcal{B}$ of $\Omega_{\overline{v},\overline{w}}(\overline{w})$ into finite parts, and a $\mathbb{P}_p$ preserving bijection between $\mathcal{A}$ and $\mathcal{B}$. We do not know if such a partition exists.

Problem 26. *Describe a large class of pairs $(v, w) \notin \mathcal{R}$ that exhibit odd symmetry, and for which there exist explicit bijections that explain the symmetry.*

Here 'large' means that the number of such pairs should be $\Theta(4^n)$. A computer search suggests that there are indeed many such pairs with odd symmetry.

The explicit bijections for even and constant symmetry appearing in Section 5 suggest many further avenues. In computer-assisted searches, we discovered many pairs that exibit even and constant symmetry, and which resemble the bijections $\varphi_{a,b}$ arising from the class $\mathcal{E}$. Specifically, pairs of the form $(ab, \bar{b}a)$ and $(a\bar{a}, b\bar{b})$ (and similar constructions) exhibit these symmetries, but constructions in the same vein as the $\varphi_{a,b}$ maps do not yield bijections in these cases. Overall, it seems there is more combinatorial structure lurking here.

Problem 27. *Describe a large class of pairs $(v, w) \notin \mathcal{R} \cup \mathcal{E}$ that exhibit even or constant symmetry, and construct explicit bijections to explain the symmetry.*

Here 'large' should mean that the number of pairs of length n is $\sim n^{-\frac{1}{2}}4^{-n}$. Here the $\sqrt{n}$ factor arises because of Theorem 9, which forces $||v||_1 = ||w||_1$ to have even or constant symmetry, and the density of such pairs is roughly $n^{-\frac{1}{2}}$.

References

[1] D. Felix. Optimal penney ante strategy via correlation polynomial identities. *the electronic journal of combinatorics*, pages R35–R35, 2006.

[2] W. Feller. *An introduction to probability theory and its applications, Volume 2*, volume 81. John Wiley & Sons, 1991.

[3] H. U. Gerber and S.-Y. R. Li. The occurrence of sequence patterns in repeated experiments and hitting times in a markov chain. *Stochastic Processes and their Applications*, 11(1):101–108, 1981.

[4] L. Guibas and A. Odlyzko. String overlaps, pattern matching, and nontransitive games. *Journal of Combinatorial Theory, Series A*, 30(2):183–208, 1981.

[5] L. J. Guibas and A. M. Odlyzko. Periods in strings. *Journal of Combinatorial Theory, Series A*, 30(1):19–42, 1981.

[6] D. A. Levin and Y. Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

[7] S.-Y. R. Li. A Martingale Approach to the Study of Occurrence of Sequence Patterns in Repeated Experiments. *The Annals of Probability*, 8(6):1171 – 1176, 1980.

[8] R. Phillips and A. J. Hildebrand. The number of optimal strategies in the penney-ante game, 2021.