Word length, bias and bijections in Penney's ante

Mathew Drexel[a]   Xuanshan Peng[a]   Jacob Richey[b]

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 This setup, commonly known as Penney's ante, was popularized by Conway, who found (in unpublished work) a simple formula for the probability that a given word occurs first. We study these 'win' probabilities in Penney's ante from an analytic and combinatorial perspective, building on previous results for the case $p = \frac{1}{2}$ and words of the same length. For words of arbitrary lengths, our results bound how large the win probability can be for the longer word. When $p = \frac{1}{2}$ we characterize when a longer word can be statistically favorable, and for $p \neq \frac{1}{2}$ we present a conjecture describing the optimal pairs, which is supported by computer computations. Additionally, we find that Penney's ante often exhibits symmetry under the transformation $p \to 1 - p$. We construct new explicit bijections that account for these symmetries, under conditions that can be easily verified by examining auto- and cross-correlations of the words.

Mathematics Subject Classifications: 60C05, 05A19

1 Introduction

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

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

If word v occurred first, i.e. $\tau_v < \tau_w$, 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, which are approximately geometric, appears in Feller's classic probability text [3]. Conway popularized the 'race'

---

version as an example of a non-transitive tournament (on the set of possible words), and gave an explicit formula for the odds that one word occurs before another. The formula, summarized in Theorem 4, arises from the so-called *Abracadabra* martingale construction, which is folklore among probabilists (see for example Section 17.3.2 of [10]). Conway's formula was generalized by Guibas/Odlyzko and Li [7][8], who found linear systems describing the win probabilities and relevant generating functions for a race among an arbitrary (finite) family of words, and later to underlying sequences generated from an arbitrary finite-state Markov chain [13][5].

Our aim in this work is to give a birds eye view of the combinatorial structure of Penney's ante through the lens of the win probability $\mathbb{P}_p(\tau_v < \tau_w)$, where $\mathbb{P}_p$ denotes iid Bernoulli(p) measure. The overarching questions we have in mind are: for fixed $p \in (0, 1)$ and pairs of words (v, w) satisfying some natural conditions, can we describe the maximum or minimum 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, and when can those symmetries be explained by a bijection?

For the unbiased case, previous works have identified the optimal choice among words of the same length and described the optimizers [12][1]. Results are also available in the case where the underlying digit sequence is iid from a distribution with finite support [15] [4]. We extend these results to words of different lengths, focusing on the question: when can a longer word possibly beat a shorter one? We prove that at $p = \frac{1}{2}$, a longer word can be a favorable choice only in a select few cases. 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 (see Theorem 1.6 of [7]). In the case $p \neq \frac{1}{2}$, we give a conjecture along similar lines, which is supported by computer computations. Interestingly, in this case the optimum win probability appears to occur for words that are essentially periodic with period 1 or 2, depending on the value of p. See Sections 4.1 and 6 for further discussion.

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 the extreme case where p is close to 0, we show in Theorem 12 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 small subtlety). One way to formalize this heuristic is the following conjecture, which is supported by computer computations (see Figure 1):

Conjecture 1. For any 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 as a function of $p \in \left(0, \frac{1}{2}\right)$.

A natural guess for how to prove Conjecture 1 is to show that for each fixed pair (v, w), the win probability $\mathbb{P}_p(\tau_v < \tau_w)$ is a monotone (or constant) function of $p \in \left(0, \frac{1}{2}\right)$. Surprisingly, this is false: a small example is the pair $(v, w) = (100010, 001100)$, with win probability $\mathbb{P}_p(\tau_v < \tau_w) = \frac{1}{2}\left(1 - p(1 - p)^4\right)$ taking its minimum value at $p = \frac{1}{5}$. See Section 6 for further discussion.
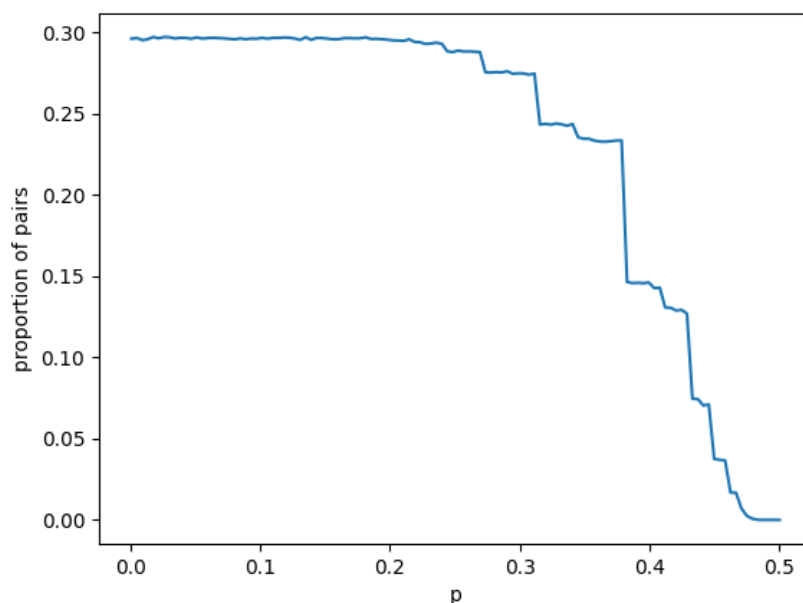
Figure 1: The approximate proportion of pairs of length at most $n = 17$ where the longer word is statistically favorable over the shorter word, plotted for $p \in (0, \frac{1}{2})$, as in Conjecture 1.

For certain pairs (v, w), the win probability exhibits some symmetry under the transformation $p \to 1 - p$, i.e. exchanging 0s and 1s. The most obvious way this can occur is if the word w is obtained from v by exchanging 0s and 1s, in which case an explicit bijection (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 (where $\overline{u}$ is the bit-flip of u).

Using Conway's formula (equation (6)), one can always express the win probability $\mathbb{P}_p(\tau_v < \tau_w)$ as a rational function of p, and explicitly check whether some symmetry holds. However, we find that many such symmetries occur for non-obvious reasons; and crucially, since the proof of Conway's formula relies (at its root) on the optional stopping theorem for martingales, it does not intrinsically give bijective proofs of these kinds of identities. This leads us to study the underlying combinatorial structure directly: namely, we consider the set Ω of all finite binary words that avoid the words v and w, and seek explicit bijections between certain subsets of Ω. Specifically, we will seek bijections φ between sets of such words that either preserve the iid measure $\mathbb{P}_p$ or exchange $\mathbb{P}_p$ with $\mathbb{P}_{1-p}$, and are locally computable: that is, to determine a given digit of the image word $\varphi(x)$, one should only have to check a bounded-sized window of x, or a window of finite expected size. This is similar to the approach taken by the authors of [2], who show (by

exact counting) a symmetry for the pair $(111, 101)$ and a specific value of p. In Proposition 13 and Theorem 15 we present large families of pairs for which we construct new, explicit, local bijections. We also find an explicit bijection for the pair $(v, w) = (1100, 1010)$ which arises from a natural graph construction, and demonstrates for any $p \in (0, 1)$ the equality $\mathbb{P}_p(\tau_v < \tau_w) = \mathbb{P}_{1-p}(\tau_v < \tau_w)$ (both are equal to $\frac{1+p(1-p)}{2+p(1-p)}$).

Remark: Our results on bijections are tangential to *symbolic dynamics*, a field at the intersection of combinatorics, dynamics, and ergodic theory – see [9]. The central objects in symbolic dynamics are *shift spaces*, akin to the sets Ω, and the relevant isomorphisms are *conjugacies*, akin to the maps φ. We note that the condition appearing in Theorem 13 gives a class of simple conjugacies, while the more interesting bijections constructed in the proof of Lemma 20 give bijections between the corresponding shift spaces with *finite expected coding-radius* (see the Remark at the end of Section 3). Typically in symbolic dynamics one studies the *measure(s) 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 (except when $p = \frac{1}{2}$).

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 (6), for the win probability. In Section 3 we advertise our main results and their implications in detail. Section 4 covers the results and conjectures regarding words of different lengths, and Section 5 presents bijections for fixed words of the same length. In Section 6 we discuss some conjectures, questions, and possible future directions.

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 bit-flip 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 bit-flip of v, i.e. $\overline{v} = \overline{v_1 v_2} \cdots \overline{v_n}$. 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 x of the form $x = uv$ for some (possibly empty) word u such that v appears exactly once as a subword of x (as the suffix), and w does not occur as a subword of x. Similarly, let $\Omega_{v,w}(w)$ be the set of binary words $x = uw$ such that w appears exactly once as a subword of x and v does not appear as a subword of x. Additionally, let $\Omega_{v,w} = \Omega_{v,w}(v) \cup \Omega_{v,w}(w)$, $\Omega^*_{v,w}(v) = \{u : uv \in \Omega_{v,w}(v)\}$, and $\Omega^*_{v,w}(w) = \{u : uw \in \Omega_{v,w}(w)\}$.

$\Omega_{v,w}$ is exactly the set of sequences one can observe in Penney's ante, since if an instance of the game produces a binary sequence x, then exactly one copy of exactly one of the words v, w occurs in x, as the suffix. The set of sequences where v wins is $\Omega_{v,w}(v)$,

and the set of sequences where w wins is $\Omega_{v,w}(w)$. Note that $\Omega_{v,w}$ is symmetric in v, w, but that in general the sets $\Omega^*_{v,w}(v)$ and $\Omega^*_{v,w}(w)$ may be different.

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:

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

Also, let $\overline{\mathcal{O}}(v,w)$ be the set of integers r such that there exists a word u of length r where $u \in \mathcal{O}(v,w)$, i.e.

$$\overline{\mathcal{O}}(v,w) = \{r \in \{1, 2, \ldots\} : \exists u \in \mathcal{O}(v,w), |u| = r\}. \tag{3}$$

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

$$vw_p = \sum_{u \in \mathcal{O}(v,w)} \mathbb{P}_p(u)^{-1} = \sum_{u \in \mathcal{O}(v,w)} p^{-||u||_1}(1-p)^{-||u||_0}. \tag{4}$$

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 polynomial pops out of a clever martingale construction, sometimes called the *Abracadabra* martingale [14][8]. Where it does not cause confusion we abuse notation and use xy for concatenation of words, and xy_p for the correlation polynomial value, since both are common in the literature.

Our main tool is the following formula, originally derived by Conway in unpublished work (see [8] for detailed proofs).

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

$$\mathbb{E}_p \tau_v = vv_p, \tag{5}$$

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

$$\mathrm{Win}(v, w; p) := \mathbb{P}_p(\tau_v < \tau_w) = \mathbb{P}_p(\Omega_{v,w}(v)) = \frac{ww_p - wv_p}{ww_p + vv_p - wv_p - vw_p}. \tag{6}$$

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 [6]). *Let v be a binary word of length n. The set*

$$\{n - i : v_1 v_2 \cdots v_i \in \mathcal{O}(v,v)\} \tag{7}$$

is closed under addition.

Additionally, in Section 5 we will use the following 'bad prefix' sets:

Definition 6. For words v, w, let $D(v, w)$ denote the set of non-empty prefixes of v corresponding to non-trivial overlaps of v, w: that is, $D(v, w)$ 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 $F(v, w) := D(v, v) \cup D(w, w) \cup D(w, v) \cup D(v, w)$.

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

Example 7. Consider the pair $(v, w) = (1101110, 0110)$. We have

$$\mathcal{O}(v, v) = \{1101110, 110\}, \ \ \mathcal{O}(w, w) = \{0110, 0\}, \ \ \mathcal{O}(v, w) = \{0\}, \ \ \mathcal{O}(w, v) = \{110\}, \quad (8)$$

which gives

$$vv_p = p^{-5}(1-p)^{-2} + p^{-2}(1-p)^{-1}, \ \ ww_p = p^{-2}(1-p)^{-2} + (1-p)^{-1}, \quad (9)$$

$$vw_p = (1-p)^{-1}, \ \ wv_p = p^{-2}(1-p)^{-1}, \ \ \mathrm{Win}(v, w; p) = \frac{p^4(1 + p - p^2)}{1 + p^3}. \quad (10)$$

Similarly,

$$D(v, v) = \{1101\}, \ \ D(w, w) = \{011\}, \ \ D(v, w) = \{110111\}, \ \ D(w, v) = \{0\}, \quad (11)$$

and $F(v, w) = \{0, 011, 1101, 110111\}$.

3   Overview & Results

Recall that for $p \in (0, 1)$, $\mathbb{P}_p(\tau_v < \tau_w) = \mathrm{Win}(v, w; p)$ is the probability that v beats w in Penney's ante. Our first main result, proved in Section 4, characterizes when a longer word can be statistically favorable over 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 $\mathrm{Win}\left(v, w; \frac{1}{2}\right) \geqslant \frac{1}{2}$, then $w \in \{1^n, 0^n\}$.*

Since by Theorem 4 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. We also calculate the optimal win probability and the optimizing pair for a word that is k digits longer than its opponent:

Theorem 9. *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*

$$\mathrm{Win}\left(v, w; \frac{1}{2}\right) < \frac{2}{1 + 2^k} \quad (12)$$

and equality is achieved asymptotically as $n \to \infty$ for the pair $(v, w) = (0^{k+1}1^{n-1}, 1^n)$.

It follows immediately that for $p = \frac{1}{2}$ and $k \geqslant 2$, a word of length $n + k$ can never have favorable odds against a word of length n. A well-known rule 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 [7]. Heuristically, this is a good choice because sometimes, when w is about to occur, v has just occurred. We note that the proof of Theorem 8 and the explicit pair appearing in Theorem 9 show that this rule of thumb holds for the case $k \geqslant 1$.

The case $p \neq \frac{1}{2}$ is qualitatively different from $p = \frac{1}{2}$ in the following way. At $p = \frac{1}{2}$, the probability of a word depends only on its length, but this is false when $p \neq \frac{1}{2}$. Thus, words with higher probability under $\mathbb{P}_p$ will have an advantage in Penney's ante, regardless of their length. 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 approximately achieve this is to require v and w to have the same number of 1s. Thus we seek to optimize over the following set:

Definition 10. For any integer $n \geqslant 2$ and $k \geqslant 1$, let $W(n, k)$ denote the set of pairs (v, w) of words of length $n + k$ and n respectively such that $||v||_1 = ||w||_1$ and w is not a subword of v.

Computer calculations suggest that the optimal pairs in $W(n, k)$ can be identified as follows. We discuss this conjecture and its implications further in Section 4.1.

Conjecture 11. For $k \in \{1, 2, \ldots\}$ and n sufficiently large, there exists $r = r(n, k)$ such that:

- If $k = 1$, then

$$\underset{(v,w) \in W(n,k)}{\arg\max} \text{Win}(v, w; p) = \begin{cases} (001^{n-1}, 1^{n-1}0), & 0 < p \leqslant r(n, 1) \\ (101^{n-1}, 1^n), & r(n, 1) \leqslant p < 1 \end{cases} \tag{13}$$

- If $k \geqslant 2$, then

$$\underset{(v,w) \in W(n,k)}{\arg\max} \text{Win}(v, w; p) = \begin{cases} (0^{k+1}(10)^{m-1}1, (10)^m), & 0 < p \leqslant r(n, k) \text{ and } n = 2m \\ (0^{k+1}(01)^m, (01)^m0), & 0 < p \leqslant r(n, k) \text{ and } n = 2m + 1 \\ (0^{k-1}101^{n-1}, 1^n), & r(n, k) < p < \frac{1}{2} \end{cases} \tag{14}$$

  where m is an integer.

We now turn to properties of the win probability, viewed as a function of $p \in (0, 1)$. For a fixed pair v, w of words with the same length, we show that the limit $p \to 0$ can only give three possible win probabilities:

Proposition 12. *For any words v, w of the same length, $\lim_{p \to 0} \text{Win}(v, w; p) \in \{0, \frac{1}{2}, 1\}$. 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 12 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 12, constant symmetry occurs if and only if $\mathrm{Win}(v, w; p) = \frac{1}{2}$ for all $p \in (0, 1)$ (if $v \neq w$). The pair $(1100, 1010)$ is an example with even but not constant symmetry, which we analyze in detail in Section 5.3. While these symmetries can occur for trivial reasons – for example, $w = \bar{v}$ easily implies odd symmetry – most of the time it is not obvious why we observe such a symmetry. Our main results, Theorems 13 and 15, identify large classes of pairs where these symmetries occur, and give explicit bijections which explain the symmetry.

Theorem 13. *Let $\mathcal{R}$ denote the family of pairs of binary words (v, w) of the same length such that $\Omega^*_{v,w}(v) = \Omega^*_{v,w}(w)$ (see Definition 2).*

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

$$\mathrm{Win}(v, w; p) = \frac{\mathbb{P}_p(v)}{\mathbb{P}_p(v) + \mathbb{P}_p(w)} = \frac{p^s}{p^s + (1-p)^s}, \tag{15}$$

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 $\mathcal{R}$ holds for an asymptotically positive fraction of pairs, in the following sense:

Conjecture 14. Let $\mathcal{R}_n$ denote the sub-family of pairs in $\mathcal{R}$ of length n. The limit

$$\lim_{n \to \infty} 4^{-n} |\mathcal{R}_n| \tag{16}$$

exists and is strictly positive.

It is easy to show that if v and w have trivial self and cross overlap sets, i.e. if $\mathcal{O}(v, w) = \mathcal{O}(w, v) = \emptyset$, $\mathcal{O}(w, w) = \{w\}$ and $\mathcal{O}(v, v) = \{v\}$, then $(v, w) \in \mathcal{R}$, and equation (6) gives the conclusions *i* and *ii*. However, an exact computer search with $n = 15$ returned the approximate value .15753 for the limit in Conjecture 14, while the proportion of pairs that have trivial auto- and cross- overlap sets is approximately .02309. Thus, Theorem 13

gives a much stronger result than what one obtains by a naive application of equation (6). The pair $v = 000100, w = 001110$ is an example where property $\mathcal{R}$ holds – see Lemma 19 – but the auto- and cross-correlations are non-trivial.

Still, property $\mathcal{R}$ is not necessary to have odd symmetry. The minimal example is 1000 and 0110: this pair has odd symmetry, with $\text{Win}(1000, 0110; p) = 1 - p$ – which is what one would obtain by applying equation (15) to (v, w) – 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 bijective argument that explains the odd symmetry in this case. See Section 6 for further discussion.

Similarly, property $\mathcal{R}$ is not necessary for constant symmetry. 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, novel bijection in Section 5, and generalize that construction to obtain a large family of such pairs in the following Theorem.

Theorem 15. *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 appearing in Theorem 15 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. Although similar bijections likely work to explain even symmetry for other pairs, we do not know how to generalize this bijection 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. Recall the correlation polynomials (Definition 3), and for simplicity, 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,

$$\text{Win}(v, w; p) < \frac{1}{2} \iff ww - wv < vv - vw. \tag{17}$$

Note that $n - 1 \notin \overline{\mathcal{O}}(w, w)$, since otherwise by Theorem 5, $\overline{\mathcal{O}}(w, w) = \{1, 2, \ldots, n - 1, n\}$, 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 (17) in each one.

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

$$vw \leqslant \sum_{1 \leqslant i \leqslant n-2} 2^i = \frac{1}{2} 2^n. \tag{18}$$

If $n - 2 \in \overline{\mathcal{O}}(w, w)$, then $n - x \notin \overline{\mathcal{O}}(w, w)$ 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

$$ww \leqslant \left(\sum_{\substack{1 \leqslant i \leqslant n \\ i \text{ even}}} 2^i \right) \wedge \left(\sum_{\substack{1 \leqslant i \leqslant n \\ i \neq n-2, n-1}} 2^i \right) \leqslant \frac{4}{3} 2^n. \tag{19}$$

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

$$ww - wv \leqslant \frac{4}{3} 2^n \text{ and } vv - vw \geqslant 2^{n+1} - \frac{1}{2} 2^n = \frac{3}{2} 2^n. \tag{20}$$

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

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

$$v = v_1 v_2 w_1 w_2 \cdots w_{n-1} = v_1 v_2 v_3 w_1 w_2 \cdots w_{n-2} \text{ and thus } w = w_1 w_1 \cdots w_1 w_n. \tag{21}$$

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, $n \notin \overline{\mathcal{O}}(v, w)$, so $vw \leqslant \sum_{i=1}^{n-1} 2^i \leqslant 2^n$. Therefore

$$ww - wv \leqslant 2^n - 0 = 2^n, \text{ and } vv - vw \geqslant 2^{n+1} - 2^n = 2^n, \tag{22}$$

as desired.

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

$$vw \leqslant \sum_{\substack{1 \leqslant i \leqslant n-1 \\ i \neq n-2}} 2^i \leqslant \frac{3}{4} 2^n, \tag{23}$$

and

$$ww \leqslant \sum_{\substack{1 \leqslant i \leqslant n \\ i \neq n-2, n-1}} 2^i \leqslant 2^n + 2^{n-2} = \frac{5}{4} 2^n. \tag{24}$$

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

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

$$w = \begin{cases} (10)^{\frac{n}{2}}, & n \text{ even} \\ (10)^{\frac{n-1}{2}} 1, & n \text{ odd} \end{cases} \tag{25}$$

and

$$v = \begin{cases} v_1 v_2 (10)^{\frac{n}{2}-1} 1, & n \text{ even} \\ v_1 v_2 (10)^{\frac{n-1}{2}}, & n \text{ odd} \end{cases} \tag{26}$$

In both cases, by a direct computation,

$$vw = \sum_{\substack{1 \leqslant i \leqslant n-2 \\ i \text{ odd}}} 2^{n-i} \leqslant \frac{2}{3} 2^n \tag{27}$$

and

$$ww = \sum_{\substack{0 \leqslant i \leqslant n-1 \\ i \text{ even}}} 2^{n-i} \leqslant \frac{4}{3} 2^n. \tag{28}$$

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

Theorem 8 is tight, in the sense that the inequality Win $\geqslant \frac{1}{2}$ cannot be replaced by Win $\geqslant \frac{1}{2} - \epsilon$ 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.)

We now prove Theorem 9, which gives a tight upper bound for the win probability in terms of the length difference k between v and w.

Proof of Theorem 9. We only need crude bounds on each of the overlap polynomials vv, ww, wv, vw: namely, $vv \geqslant 2^{n+k}$, $ww \leqslant \sum_{i=1}^{n} 2^i \leqslant 2^{n+1}$, $wv \geqslant 0$, and since v is not a subword of w, $n \notin \overline{\mathcal{O}}(v, w)$, so $vw \leqslant \sum_{i=1}^{n-1} 2^i \leqslant 2^n$. Thus, by Theorem 4,

$$\mathrm{Win}\left(v, w; \frac{1}{2}\right) \leqslant \frac{2^{n+1}}{2^{n+k} + 2^{n+1} - 2^n} = \frac{2}{1 + 2^k}. \tag{29}$$

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. $\qquad\square$

4.1 Long vs short with bias

We verified Conjecture 11 by an exact computer search for all pairs of words with $n \leqslant 11$ and $k \leqslant 4$. Note that all the optimal pairs appearing there have the property that v has $w_1 w_2 \cdots w_{n-1}$ as a suffix, following the rule of thumb for the optimal choices in the case $p = \frac{1}{2}$ and $k = 0$. It is surprising that only words that are periodic (or contain a large, periodic subword) 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 could not find a proof along the same lines as the proof of Theorem 9. 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 11 holds, and recall the notation there. Fix a pair of binary words $(v, w) \in W(n, k)$.*

- *For $k = 1$,*

$$\mathrm{Win}(v, w; p) \leqslant \begin{cases} 1 - p, & 0 < p \leqslant r(n, 1) \\ \frac{1}{2 - p}, & r(n, 1) \leqslant p < \frac{1}{2} \end{cases} \tag{30}$$

The threshold satisfies

$$r(n, 1) = \kappa + O(\kappa^n) \text{ as } n \to \infty, \text{ where } \kappa = \frac{3 - \sqrt{5}}{2}. \tag{31}$$

- *For $k \geqslant 2$,*

$$\mathrm{Win}(v, w; p) \leqslant \begin{cases} \frac{(1-p)^k}{1 - p(1 - (1-p)^k) + p^2}, & 0 < p \leqslant r(n, k) \\ \frac{(1-p)^{k-1}}{1 + (1-p)^k}, & r(n, k) \leqslant p < \frac{1}{2} \end{cases} \tag{32}$$

The threshold $r(n, k)$ is, up to $o(1)$ terms as $n \to \infty$, the unique root in $(0, \frac{1}{2})$ of the polynomial $(1 - 2z)(1 - z)^k - z^2$. In particular,

$$r(n, k) \approx \frac{k + 2 - \sqrt{k^2 - 4k + 8}}{4k - 2} \text{ as } n, k \to \infty. \tag{33}$$

We omit the proofs, which are straightforward calculations applying Theorem 4 to the explicit pairs appearing in Conjecture 11. The upper bounds in Corollary 16 are continuous at the transition point $r(n, k)$. For the case where $k \geqslant 2$ and n is odd, the optimal pair $(0^{k+1}(01)^m, (01)^m 0)$ 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' vs w' race under $\mathbb{P}_{p'}$? Here we restrict our attention to the case $(v', w') = (v, w)$ and $p' = 1 - p$. 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, which we re-state here (in a slightly different, but equivalent, form):

Definition 17. Let v, w be any binary words with the same length. We say a pair (v, w) has **odd symmetry** if $\mathrm{Win}(v, w; 1 - p) = 1 - \mathrm{Win}(v, w; p)$ for all $p \in (0, 1)$; **even symmetry**, if $\mathrm{Win}(v, w; 1 - p) = \mathrm{Win}(v, w; p)$ for all $p \in (0, 1)$; 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 12.

Proof of Proposition 12. We start with the first conclusion. The result is clear if either v or w contains the other as a subword, so assume this is not the case. Let $r = \max\{||v||_1, ||w||_1\}$. 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 for non-negative integer k. Also, at most one of vw_p or wv_p may contain a term of the form $p^{-r}(1 - p)^{-j}$ for non-negative integer j: indeed, if (without loss of generality) vw_p contains a term of the form $p^{-r}(1 - p)^{-j}$, say $u \in \mathcal{O}(v, w)$ with $||u||_1 = r$, then we must have $v = 0^a 1u'0^b$ and $w = 0^c 1u'0^d$ for some non-negative integers a, b, c, d with $c \leqslant a$, and where u' is a subword of u. Since 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

$$\mathrm{Win}(v, w; p) = \frac{\sum_{0 \leqslant i,j \leqslant m}(\alpha_{ij} - \beta_{ij})p^{-i}(1 - p)^{-j}}{\sum_{0 \leqslant i,j \leqslant m}(\alpha_{ij} - \beta_{ij} + \gamma_{ij} - \delta_{ij})p^{-i}(1 - p)^{-j}}, \tag{34}$$

where the $\alpha, \beta, \gamma, \delta$ coefficients are $\{0, 1\}$ valued. Let r' denote the maximum power of p^{-1} that appears in either sum (after any cancellations). By the observations in the previous paragraph, $r' = r$, 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 $||v||_1 \neq ||w||_1$, 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 $\mathcal{O}(w, v)$ contains a word with 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 that we say *odd symmetry* holds if

$$\text{Win}(v, w; p) = 1 - \text{Win}(v, w; 1 - p) \text{ for all } p \in (0, 1), \tag{35}$$

or equivalently, if

$$\mathbb{P}_p(\tau_v < \tau_w) = 1 - \mathbb{P}_{1-p}(\tau_v < \tau_w) = \mathbb{P}_{1-p}(\tau_w < \tau_v) = \mathbb{P}_p(\tau_{\overline{w}} < \tau_{\overline{v}}). \tag{36}$$

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 $\mathbb{P}_p(\tau_v < \tau_{\overline{v}}) = \mathbb{P}_{1-p}(\tau_{\overline{v}} < \tau_v)$. A much larger (and more interesting) class of pairs with odd symmetry is the following.

Definition 18 (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 F (see Definition 6):

Lemma 19. *A pair of words (v, w) has property $\mathcal{R}$ if and only if $fv, fw \notin \Omega_{v,w}$ for all $f \in F(v, w)$.*

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

$$\Omega^*_{v,w}(v) = \Omega^*_{v,w}(w) \iff \text{ for all } x \in \Omega^*_{v,w}, (xv \notin \Omega_{v,w} \iff xw \notin \Omega_{v,w}) \tag{37}$$

$$\iff \text{ for all } x \in \{0, 1\}^{n-1}, (xv \notin \Omega_{v,w} \iff xw \notin \Omega_{v,w}) \tag{38}$$

$$\iff \text{ for all } f \in F(v, w), (fv \notin \Omega_{v,w} \iff fw \notin \Omega_{v,w}) \tag{39}$$

$$\iff \text{ for all } f \in F(v, w), (fv, fw \notin \Omega_{v,w}) \tag{40}$$

The equivalence (37) is clear. For (38), 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 (39), 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 $D(v,v) \cup D(w,v)$, and similarly for xw. The final equivalence follows from the fact that $fv, fw \in \Omega_{v,w}$ for any $f \in \{0,1\}^{n-1} \setminus F(v,w)$. $\qquad\square$

(Proof of Theorem 13). By property $\mathcal{R}$, let $\Omega^* = \Omega_{v,w}^*(v) = \Omega_{v,w}^*(w)$. By the definitions of $\mathrm{Win}(v,w;p)$ and $\Omega_{v,w}^*(v)$ and using that $\mathbb{P}_p$ is iid,

$$\mathrm{Win}(v,w;p) = \mathbb{P}_p(\Omega_{v,w}(v)) = \mathbb{P}_p(\Omega^*)\mathbb{P}_p(v), \tag{41}$$

and similarly,

$$\mathbb{P}_p(\Omega_{v,w}(w)) = \mathbb{P}_p(\Omega^*)\mathbb{P}_p(w). \tag{42}$$

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

$$\mathrm{Win}(v,w;p) = \frac{\mathbb{P}_p(\Omega^*)\mathbb{P}_p(v)}{\mathbb{P}_p(\Omega^*)\mathbb{P}_p(v) + \mathbb{P}_p(\Omega^*)\mathbb{P}_p(w)} = \frac{\mathbb{P}_p(v)}{\mathbb{P}_p(v) + \mathbb{P}_p(w)}. \tag{43}$$

Straightforward algebra gives the last equality of i, which directly implies odd symmetry, and constant symmetry if $\mathbb{P}_p(v) = \mathbb{P}_p(w)$, and thus ii holds. Since the set $F(v,w)$ can be determined by checking equality of linearly many overlaps with respect to the length of v (= length of w), and similarly for a condition of the form $fv \in \Omega_{v,w}$ or $fw \in \Omega_{v,w}$ where f has length at most that of v, part iii follows from Lemma 19. $\qquad\square$

5.2 Constant symmetry

Recall that we say a pair v, w has *constant symmetry* if $\mathrm{Win}(v,w;p) = \frac{1}{2}$ for all $p \in (0,1)$. Note that Theorem 13 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 such example, which we will then generalize to a large family of such examples in Theorem 15.

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 occurrences 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 $z = xa^j v \in \Omega_{v,w}(v)$ where $j \geqslant 0$ is maximal. Write $j = j(z)$ for the maximal such j. Then define the map $\varphi_{a,b}$ by

$$\varphi_{a,b}(xa^j v) = xb^j w, \tag{44}$$

and define $\varphi_{b,a}$ similarly, with the roles of a and b reversed. Note that $\mathbb{P}_p(a) = \mathbb{P}_p(b)$, and thus $\mathbb{P}_p(v) = \mathbb{P}_p(w)$, so both $\varphi_{a,b}$ and $\varphi_{b,a}$ are measure-preserving. Also, for $z = xa^{j(z)}v \in \Omega_{v,w}(v)$, x cannot not have b as a suffix (else z would contain a copy of w), and thus $j(\varphi_{a,b}(z)) = j(xb^{j(z)}w) = j(z)$. (And the corresponding fact holds for the reverse map.) Then, assuming that $\varphi_{a,b}$ maps $\Omega_{v,w}(v)$ to $\Omega_{v,w}(w)$ (and vice versa for $\varphi_{b,a}$) – this is the crux of the argument and will be proved below – we would have that $\varphi_{a,b} = \varphi_{b,a}^{-1}$, so $\varphi_{a,b}$ is a measure-preserving bijection, completing the proof of the first part of the Lemma. We will show that $\varphi_{a,b}$ has image in $\Omega_{v,w}(w)$; the details for $\varphi_{b,a}$ are essentially the same.

Write $\Omega_{v,w} = \Omega$ for short, and for a word u, let $u_{[i]} = u_1 u_2 \cdots u_i$ denote the i prefix of u. We must 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,

$$yv \in \Omega(v) \iff y \neq y'u \text{ for } u \in \mathcal{D}(v,v) \cup \mathcal{D}(w,v) = \{b, v_{[6]}, w_{[7]}\} \tag{45}$$

and

$$yw \in \Omega(w) \iff y \neq y'u \text{ for } u \in \mathcal{D}(w,w) \cup \mathcal{D}(v,w) = \{a, v_{[6]}, w_{[7]}\}. \tag{46}$$

Consider a word $z = xa^j v \in \Omega(v)$ with $j(z) = 0$, i.e. a word $xv \in \Omega(v)$ where x does not end in a. By statement (45), x does not end in $a, b, v_{[6]}$ or $w_{[7]}$; and then by statement (46), $xw \in \Omega(w)$ 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 $z = xa^j v \in \Omega(v)$ with $j(z) \geqslant 1$. Since j is maximal, by statement (45), x does not end in $a, b, v_{[6]}$ or $w_{[7]}$. Observe that

$$\mathcal{D}(v, bb) = \{a, v_{[6]}\} \text{ and } \mathcal{D}(w, bb) = w_{[7]}. \tag{47}$$

Since we assumed x doesn't end in any of these words, we obtain that $xb^j w = xb^{j+1}a$ 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 $xb^j w \in \Omega(w)$, as desired.

Since $\mathbb{P}_p(v) = \mathbb{P}_p(w)$, the existence of $\varphi_{a,b}$ implies (by the same calculation as in the proof of Proposition 13) that $\mathrm{Win}(v, w; p) = \frac{1}{2}$. $\qquad\square$

The construction in Lemma 20 generalizes 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 binary words a, b and a finite set S of binary 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 $m \geqslant 2$, $D(v,a^m) \cup D(w,a^m) \subset S \cup \{a,b\}$ and $D(v,a^m) \cup D(w,b^m) \subset S \cup \{a,b\}$

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

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

(Proof of Theorem 15). The argument in Lemma 20 adapts immediately to this setting: properties II-VI imply, by the same argument, that the map $\varphi_{a,b} : \Omega_{v,w}(v) \to \Omega_{w,v}(w)$ given by

$$\varphi_{a,b}(xv) = x'b^j w \text{ where } j = \max\{j' : a^{j'} \text{ is a suffix of } x\} \tag{48}$$

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. For example, the condition $D(v,a^m) \subset S \cup \{a,b\}$ does not depend on m if a^m has length greater than that of v, and similar facts hold for all the conditions in IV, V and VI. Thus at most polynomially many comparisons are required. Finally, an easy computation shows that for all $k \geqslant 2$, words $a = 01^k 0, b = 01^{k-1} 01$ give rise to a pair $(a_k b_k, b_k a_k) \in \mathcal{E} \setminus \mathcal{R}$. $\qquad\square$

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 in terms of a graphical representation of the space $\Omega_{v,w}$.

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 2. This construction is a common way to compute recursions, probabilities and expectations explicitly in Penney's ante [11]. (In the language of symbolic dynamics, it is similar to the follower set edge-shift graph of the corresponding shift space, see for example Chapter 3.2 of [9].) We summarize the connection between Penney's ante and this graph in the following lemma.
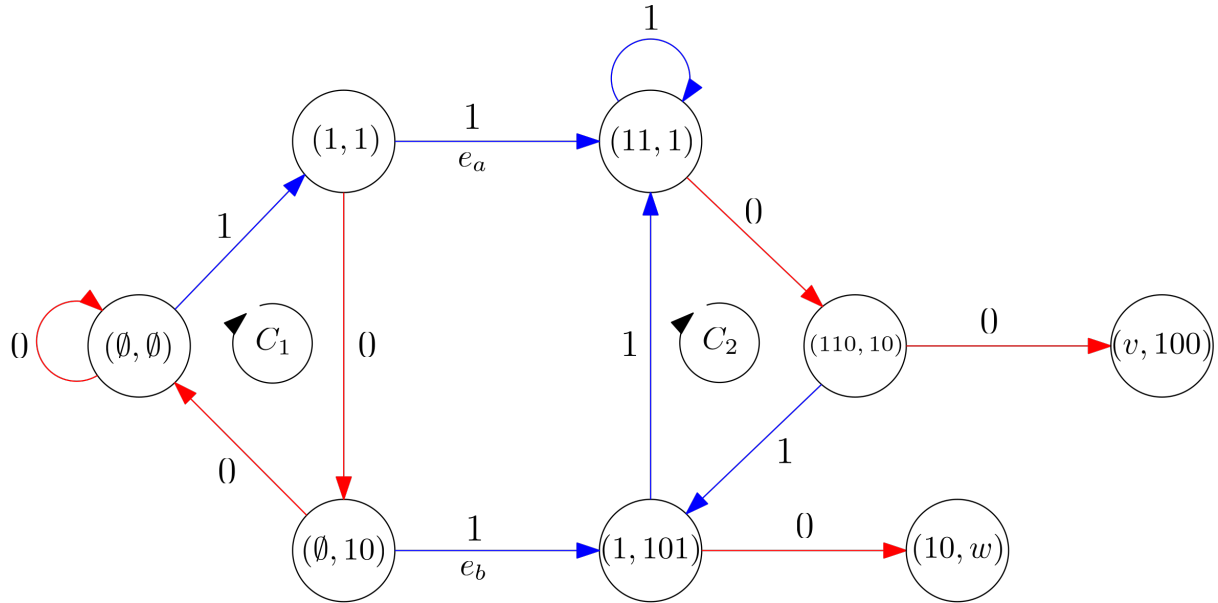
Figure 2: The labeled graph $G_{v,w}$ encoding the space $\Omega_{v,w}$, for $v = 1100, w = 1010$. Elements of $\Omega_{v,w}$ are in bijection with labeled paths in $G_{v,w}$ starting at $(\emptyset, \emptyset)$ and ending at either $(v, 100)$ or $(10, w)$, by reading the labels along the edges of the path. (Blue edges are labeled 1, and red edges are labeled 0.)

Lemma 23. *Let $G_{v,w}$ be the graph appearing in Figure 2, and denote by $\mathcal{Q}$ the set of directed edge-paths in $G_{v,w}$ starting from vertex $(\emptyset, \emptyset)$ and ending at either $(v, 100)$ or $(10, w)$. Denote by $\mathcal{Q}(v)$ (resp. $\mathcal{Q}(w)$) the subset of $\mathcal{Q}$ ending at $(v, 100)$ (resp. $(10, w)$.) Also, let $\mathcal{L}$ denote the law of the random walk, viewed as an edge-path on $G_{v,w}$, which starts from $(\emptyset, \emptyset)$, assigns probability p to blue edges and $1 - p$ to red edges, and terminates at the first hitting time of the set $\{(v, 100), (10, w)\}$. There exists a bijection $\Pi : \mathcal{Q} \to \Omega_{v,w}$ such that the following hold:*

   i. *Π restricted to $\mathcal{Q}(v)$ is a bijection onto its range $\Omega_{v,w}(v)$, and similarly for $\mathcal{Q}(w)$ and $\Omega_{v,w}(w)$.*

   ii. *$\Pi_*(\mathcal{L}) = \mathbb{P}_p$, where Π_* denotes the pushforward (of the measure $\mathcal{L}$).*

Proof. Let $\gamma = (e_1, e_2, \ldots, e_t)$ be a path in $\mathcal{Q}$, 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 $\Pi : \mathcal{Q} \to \Omega_{v,w}$ given by

$$\Pi(\gamma) = \pi(e_1) \cdots \pi(e_t) \tag{49}$$

is the desired 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, always

---
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**   ▷ Search the space of possible prefix pairs
 5: $(v', w') \leftarrow$ 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 $\mathcal{O}$
     $w'' \leftarrow \max \mathcal{O}(w'a, w)$
     if $(v'', w'') \notin$ Vertices **then**
 10:     add (v'', w'') to Vertices
      if $v'' \neq v$ **and** $w'' \neq w$ **then**:
       add (v'', w'') to Stack
     add $(v', w') \rightarrow (v'', w'')$ to Edges   ▷ This edge is labelled a

---

keeping track of the maximal prefixes of v, w that occur as suffixes of the generated word, and exploring no deeper when a copy of either v or w is generated.

We prove i for $\mathcal{Q}(v)$; the result for w is similar. Let $z \in \Omega_{v,w}(v)$. By Algorithm 1, 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 for each binary word $z = (z_1, z_2, \ldots) \in \Omega_{v,w}(v)$, there exists a unique edge-path γ in $G_{v,w}$ with $\Pi(\gamma) = z$ obtained by choosing the edges labeled $z_1, z_2, \ldots$ in succession. By algorithm 1 (or a direct check of paths of length 4 in the graph $G_{v,w}$), any edge-path in $\mathcal{Q}$ ending a sequence of edges labeled $(1, 1, 0, 0)$ ends at vertex $(v, 100)$. Thus, since z has $v = 1100$ as a suffix, the path γ ends at vertex $(v, 100)$. Thus Π^{-1} is well-defined on $\Omega_{v,w}(v)$, and Π is surjective as a map $\mathcal{Q}(v) \rightarrow \Omega_{v,w}(v)$. Moreover, Π is injective by definition, so Π is the desired bijection. The definition of Π and the fact that $\mathbb{P}_p$ is iid gives ii. $\square$

We are now ready to prove Proposition 22.

Proof. Let $\mathcal{P}(v)$ denote the set of vertex-paths $(\emptyset, \emptyset) \rightarrow (v, 100)$ in $G_{v,w}$, and for integers $j, k \geqslant 0$, 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 $\psi : \mathcal{P}(v) \rightarrow \mathcal{P}(v)$ with the property that for all $j, k \geqslant 0$, ψ restricted to $\mathcal{P}_{j,k}$ is a bijection $\psi_{j,k} : \mathcal{P}_{j,k}(v) \rightarrow \mathcal{P}_{k,j}(v)$. 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 $1 \leqslant i \leqslant N_1(x)$, 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 $1 \leqslant i \leqslant N_2(x)$, 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_a, e_b exactly once. Let $e(x) = 1$ if e_a was traversed and $e(x) = 2$ otherwise. Altogether, any path $x \in \mathcal{P}$ is uniquely described by the tuple

$$(N_1(x), N_2(x), (s_i^1(x))_{1 \leqslant i \leqslant N_1(x)}, (s_i^2(x))_{1 \leqslant i \leqslant N_2(x)}, e(x)), \tag{50}$$

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

$$(N_1(y), N_2(y), (s_i^1(y)), (s_i^2(y)), e(y)) = (N_2(x), N_1(x), (s_i^2(x)), (s_i^1(x)), e(x)). \tag{51}$$

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

$$r = 2(N_1(z) - 1) + \sum_{i=1}^{N_1(z)} s_i^1(z) + 1\{e(z) = 2\} + N_2(z) + 1, \tag{52}$$

and

$$b = N_1(z) + 1 + 1\{e(z) = 2\} + \sum_{i=1}^{N_2(z)} s_i^2(z) + 2(N_2(z) - 1). \tag{53}$$

The terms in equation (52) 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 (53).

Observe that equations (52) and (53) 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 (51) of ψ and equations (52) and (53) for both $(r, b) = (j, k)$ with path x and $(r, b) = (k, j)$ with path $y = \psi(x)$ completes the proof. $\qquad\square$

6 Discussion

Our work suggests many questions and possible 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? Can the existing results on the optimal strategy for the $p = 1/2$ case be extended to arbitrary p, or to other underlying Markov chains? Proving Conjecture 11 would be a good starting point. Alternatively,

it may be more natural 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?

Along the same line, Figure 1, which plots the volume of pairs where the longer word is favorable in terms of p, suggests that in the limit $n \to \infty$ there is a fractal set of p values where the proportion jumps. Specifically, for positive integer n, $p \in (0, \frac{1}{2})$ and $\alpha \in (0, 1)$, let $a(n, \alpha, p)$ denote the number of pairs (v, w) of length at most n where v is longer than w and $\text{Win}(v, w; p) \geqslant \alpha$, and note that the total number of pairs (v, w) of length at most n where v is longer than w is asymptotically $\frac{1}{3} 4^{n+1}$.

Question 24. Does the limit

$$\psi(\alpha, p) = \lim_{n \to \infty} 4^{-n} a(n, \alpha, p) \tag{54}$$

exist? If so, for what values of α is it a piecewise-continuous, monotone function of p? What is the structure of the set of discontinuities?

Computer simulations suggest that there is a constant $\alpha_c \approx \frac{1}{3}$ such that $\psi(p)$ is non-increasing for $\alpha < \alpha_c$ and not monotone for $\alpha > \alpha_c$, though we were not able to simulate efficiently enough to properly investigate these questions.

As noted in the introduction, $\text{Win}(v, w; p)$ 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 12 $f(0), f(1) \in \{0, \frac{1}{2}, 1\}$. What can be said about the derivative of the win probability as a function of p? 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 $\frac{d}{dp}\text{Win}(v, w; p)$ changes sign at least k times for $p \in (0, 1)$.

We also wonder if there is some combinatorial significance to the values of p where this derivative vanishes. Do values of p where $\frac{\partial}{\partial p}\text{Win}(v, w; p) = 0$ correspond to probability spaces with some special or extremal properties, or measures $\mathbb{P}_p$ that solve some variational problem over the space $\Omega_{v,w}$?

Our results for odd symmetry deal with a majority of cases, but there are still some 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 $f : \Omega_{v,w}(v) \to \Omega_{v,w}(w)$ that transforms $\mathbb{P}_p$ to $\mathbb{P}_{1-p}$: indeed, any length-preserving f must have $f(v) = w$ since these are the only allowable words of length 4, but $\mathbb{P}_p(v) \neq \mathbb{P}_{1-p}(w)$. The next best thing would be to find a pair of partitions $\mathcal{A}$ of $\Omega_{v,w}(v)$ and $\mathcal{B}$ of $\Omega_{v,w}(w)$ into finite parts, and a bijection $\varphi : \mathcal{A} \to \mathcal{B}$ with $\mathbb{P}_p(A) = \mathbb{P}_{1-p}(\varphi(A))$ for $A \in \mathcal{A}$. We do not know if such a partition exists.

Problem 26. Characterize the pairs $(v, w) \notin \mathcal{R}$ that exhibit odd symmetry, and give explicit bijections that explain the 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 exhibit even and constant symmetry, and which resemble the pairs in $\mathcal{E}$. Specifically, pairs of the form $(ab, \bar{b}a)$ and $(a\bar{a}, b\bar{b})$ (and similar constructions) often have these symmetries, but constructions in the same vein as the maps $\varphi_{a,b}$ 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 $\sim n^{-\frac{1}{2}} 4^n$. (The $\sqrt{n}$ factor arises because of Theorem 12, which forces $||v||_1 = ||w||_1$ in the case of even or constant symmetry, and the density of such pairs is roughly $n^{-\frac{1}{2}}$.)

Acknowledgements

References

[1] Csirik, János A. Optimal Strategy for the First Player in the Penney Ante Game. *Combinatorics, Probability and Computing*, 1(4): 311-321, 1992.

[2] Dababneh, Issa and Elmer, Mark and McCulloch, Ryan. A Golden Penney Game. *Journal of Integer Sequences*, 24(21.6.2), 2021.

[3] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[4] Felix, Daniel. Optimal Penney Ante Strategy via Correlation Polynomial Identities. *The Electronic Journal of Combinatorics*, #R35, 2006.

[5] Hans U. Gerber and Shuo-Yen Robert 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.

[6] Leo J. Guibas and Andrew M. Odlyzko. Periods in strings. *Journal of Combinatorial Theory, Series A*, 30(1): 19-42, 1981.

[7] Leo J. Guibas and Andrew M. Odlyzko. String overlaps, pattern matching, and nontransitive games. *Journal of Combinatorial Theory, Series A*, 30(2): 183-208, 1981.

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

[9] Douglas Lind and Brian Marcus. *An introduction to symbolic dynamics and coding.* Second edition. Cambridge University Press, Cambridge, 2021.

[10] Levin, David A. and Peres, Yuval. *Markov chains and mixing times.* American Mathematical Soc., 2017.

[11] Nickerson, Raymond S. Penney Ante: Counterintuitive probabilities in coin tossing. *The UMAP Journal*, 28(4): 503–532, 2007.

[12] Reed Phillips and A. J. Hildebrand. The Number of Optimal Strategies in the Penney-Ante Game. *Integers: Electronic Journal of Combinatorial Number Theory*, #P27(21), 2021.

[13] Pozdnyakov, Vladimir and Steele, J. Michael. *Martingale Methods for Patterns and Scan Statistics.* Birkhäuser, 2009.

[14] A. D. Solov'ev. A Combinatorial Identity and Its Application to the Problem Concerning the First Occurrence of a Rare Event. *Theory of Probability & Its Applications*, 11(2): 276-282, 1966.

[15] Stark, Dudley. First Occurrence in Pairs of Long Words: A Penney-ante Conjecture of Pevzner. *Combinatorics, Probability and Computing*, 4(3): 279-285, 1995.