

# First-past-the-post Games

Roland Backhouse

`roland.backhouse@nottingham.ac.uk`

School of Computer Science University of Nottingham, Nottingham NG8 1BB,  
England

**Abstract.** Informally, a first-past-the-post game is a (probabilistic) game where the winner is the person who predicts the event that occurs first among a set of events. Examples of first-past-the-post games include so-called block and hidden patterns and the Penney-Ante game invented by Walter Penney. We formalise the abstract notion of a first-past-the-post game, and the process of extending a probability distribution on symbols of an alphabet to the plays of a game.

Analysis of first-past-the-post games depends on a collection of simultaneous (non-linear) equations in languages. Essentially, the equations are due to Guibas and Odlyzko but they did not formulate them as equations in languages but as equations in generating functions detailing lengths of words.

Penney-Ante games are two-player games characterised by a collection of regular, prefix-free languages. For such two-player games, we show how to use the equations in languages to calculate the probability of winning. The formula generalises a formula due to John H. Conway for the original Penney-Ante game. At no point in our analysis do we use generating functions. Even so, we are able to calculate probabilities and expected values. Generating functions do appear to become necessary when higher-order cumulatives (for example, the standard deviation) are also required.

**Keywords:** algorithmic problem solving, regular language, generating function, probabilistic game, Penney-Ante, block pattern, hidden pattern

Penney-Ante is the name of a game with pennies invented by Walter Penney [Pen74]. The two-player game is interesting because it is non-transitive; the game is also used to demonstrate the use of generating functions in the calculation of probability distributions [GO81,GKP94]. Our interest in the game began as a simple, (for us) introductory exercise in probability generating functions. It has turned out to be an exercise in applying the calculational method to the analysis of the game in the general case of an arbitrary number of players — an exercise with the surprising conclusion that generating functions are not needed for the calculation of probabilities and expected values.

Analysis of the game is substantially facilitated by a collection of simultaneous (non-linear) equations between languages. In the literature, either the equations are stated without proof [GKP94] or the equations are not given explicitly but translated directly into generating functions detailing lengths of words

[GO81]. The contribution of this paper is to record a derivation of the equations and the associated probability distributions in which naming of word length and the use of generating functions is avoided.

Our derivation has several novel features. We introduce the abstract notion of a first-past-the-post game, and we formalise the process of extending a probability distribution on symbols of an alphabet to the plays of such a game (section 2). (Multi-player) Penney-Ante games and so-called block and hidden patterns [FS09] are shown to be instances of first-past-the-post games. Such games are characterised by a collection of regular, prefix-free languages. We derive a collection of simultaneous non-linear equations in these languages and use these to show how to calculate the probability of winning (section 4).

The equations are essentially the basis for the equations in generating functions derived by Guibas and Odlyzko [GO81]. The formula we derive generalises a formula due to John Horton Conway for the original two-player Penney-Ante game. Another instance is the formula due to A.D.Solov'ev [Sol66] for the expected number of coin tosses until a given (contiguous) pattern appears. Like Guibas and Odlyzko [GO81], we also consider the generalisation of Penney-Ante games to an arbitrary number of players.

We show in section 5 that the equations in languages do not have a unique solution. This is surprising and demands further investigation.

## 1 Preliminaries

We assume familiarity with the use of regular expressions to denote languages. To avoid confusion with ordinary addition, the usual symbol “ $\cup$ ” is used to denote set union, and not “ $+$ ” (as often used in regular expressions). The symbol  $\varepsilon$  denotes the empty word and  $T$  denotes a finite set (which is fixed throughout the paper). In line with other literature on the Penney-Ante game, capital letters at the beginning of the alphabet ( $A, B$ , etc.) denote words and capital letters at the end of the alphabet ( $U, V$ , etc.) denote sets of words. The elements of  $T$  are called *symbols* and sets of words are called *languages*. Symbols are denoted by lower case letters ( $a, b$ , etc.). The length of word  $A$  is denoted by  $\#A$ . Concatenation of words and of languages is denoted by juxtaposition.

For any word  $A$  different from the empty word,  $pre.A$  is the prefix of  $A$  obtained by discarding the last symbol in  $A$ . The function  $pre$  is extended to sets by the definition: for all languages  $V$ ,

$$pre.V = \{A,a : A \in T^* \wedge a \in T \wedge Aa \in V : A\} .$$

(We use the Eindhoven notation for quantifications [Bac86,GS93,Bac03]. The notation  $\{vars : rng : term\}$  abbreviates  $\langle \cup vars : rng : \{term\} \rangle$ . In conventional notation, the dummy  $a$  in the definition of  $pre$  would be existentially quantified.)

Repeated application of  $pre$  one or more times is denoted by  $pre^+$  and zero or more times by  $pre^*$ . Thus  $pre^+.V$  is the set of all proper prefixes of words in  $V$ , and  $pre^*.V$  is  $V \cup pre^+.V$ . Note that  $pre$  distributes through set union.

For calculational purposes the following property of  $pre^+$  is used. For all words  $C$  and languages  $V$ ,

$$C \in pre^+.V \equiv \{C\}T^+ \cap V \neq \emptyset .$$

## 2 First-past-the-post Games

Penney-Ante is an instance of a class of probabilistic games for which winning is characterised by the *first* occurrence of one of a set of events, and the events are words. We begin by formalising this class of games.

**Definition 1.** Suppose  $S$  is a subset of  $T^*$ . The set  $S$  is said to be a *first-past-the-post game* if

(a)  $pre^+.S \cap S = \emptyset$  .

In words, no proper prefix of a word in  $S$  is a word in  $S$ .

(b)  $pre^*.S = \{\varepsilon\} \cup (pre^+.S)T$  .

In words, appending an arbitrary symbol of the alphabet  $T$  to a proper prefix of a word in  $S$  gives a word that prefixes a word in  $S$ .

(This informal statement expresses only that the right side of the equation is included in the left side. The opposite inclusion is obvious from the definitions of  $pre^*$  and  $pre^+$ .)

A *play* of the game is an element of  $pre^*.S$ . A *complete play* of the game is an element of  $S$ . □

A play of the game can be thought of as repeatedly throwing a die with sides labelled by the elements of  $T$ . The play starts with the empty word and, as the die is thrown, the symbol that occurs is appended to the end of the play. The play is complete when the play is in  $S$ . Property (a) states that no proper prefix of a word in  $S$  is an element of  $S$ . That is, the game ends—the play is complete—immediately an element of  $S$  is recognised. Property (b) states that the plays are the empty word or arbitrary continuations of an incomplete play. It has the consequence that any throw of the die continues an incomplete play of the game. A second consequence is that  $S$  is non-empty (because the right side of the equation is a non-empty set).

*Example 1.* With  $T = \{a,b\}$ , the table below shows examples of languages and whether or not they fulfill properties (a) and (b) of definition 1.

Language	(a)	(b)
$\{a\}$	✓	×
$\{a,ab\}$	×	×
$T^k$ ( $0 \leq k$ )	✓	✓
$T^{\leq k}$ ( $0 < k$ )	×	✓
$\{a,ba,bb\}$	✓	✓
$\{b\}^*\{a\}$	✓	✓
$\{b\}^*\{a\}\{a\}^*\{b\}\{b\}^*\{a\}$	✓	✓

□

The set  $T^k$ , where  $k$  is some fixed natural number, exemplifies the set of complete plays in a first-past-the-post game. (See example 1.) It is the game where a die is thrown exactly  $k$  times.

Generally, the set  $S$  may be assumed to be split into disjoint sets each of which is owned by one of the players. When the play is complete, the owner of the play is the winner. The Penney-Ante game assumes that two players each choose one word. The reason for this assumption is that the game is then non-transitive: if one player chooses one word it is always possible for the second player to choose a word that gives a better than evens chance of winning. This, however, is not the focus of our investigation. For our purposes, the number of players can be arbitrary as can be the number of words each player chooses. There is no reason why games with fewer or more than two players should not be allowed, or why each player should choose just one word. “Games” with one player are associated with pattern-matching problems. See section 4.

We assume that the outcome of each single throw of the die is given by some probability distribution  $p$ . The outcomes of separate throws are assumed to be independent. This suggests the following definition.

**Definition 2.** Let  $p$  be a function with domain  $T$  and range the set of real numbers. We define the function  $h_p$  with domain  $T^*$  inductively by

- (a)  $h_p.\varepsilon = 1$  ,
- (b)  $h_p.Ba = h_p.B \times p.a$  , for all  $B \in T^*$  and  $a \in T$ .

The function  $h_p$  is extended to languages by defining, for all  $V$ , where  $V \subseteq T^*$ ,

$$h_p.V = \langle \Sigma A : A \in V : h_p.A \rangle .$$

The function  $e_p$  is defined on languages by, for all  $V$ , where  $V \subseteq T^*$ ,

$$e_p.V = \langle \Sigma A : A \in V : h_p.A \times \#A \rangle .$$

(Note: these definitions assume that the summations are well defined. In all the concrete examples discussed in this paper, this is indeed the case.)  $\square$

Theorem 1 shows that, if  $p$  is a probability distribution on  $T$ ,  $h_p$  is a probability distribution on a first-past-the-post game  $S$ . The value of  $e_p.S$  is then interpreted as the “expected” length of the game. It is important to note, however, that definition 2 does *not* assume that  $p$  is a probability distribution. We apply definition 2 just as often when  $p$  and/or  $h_p$  cannot be viewed as probability distributions.

Typically languages are defined syntactically — by a combination of regular expressions and equations (aka grammars). Unambiguity of syntactic definitions is useful in the evaluation of the functions  $h_p$  and  $e_p$ . This is made precise in the following definitions and lemmas.

**Definition 3 (Unambiguous Expressions).** Let  $U$  and  $V$  be expressions denoting languages  $L.U$  and  $L.V$ , respectively. We say that the expression “ $U \cup V$ ”

is *unambiguous* if  $L.U \cap L.V = \emptyset$  (i.e. the languages are disjoint). We say that the expression “ $UV$ ” is *unambiguous* if, for all words  $A, A', B$  and  $B'$ ,

$$A, A' \in L.U \wedge B, B' \in L.V \wedge AB = A'B' \Rightarrow A = A' \wedge B = B' .$$

We say that the expression “ $U^*$ ” is *unambiguous* if, for all natural numbers  $k$  and  $k'$ , and sequences of words  $A_i$  ( $1 \leq i \leq k$ ) and  $B_j$  ( $1 \leq j \leq k'$ ) all of which are elements of  $L.U$ ,

$$A_1 \dots A_k = B_1 \dots B_{k'} \Rightarrow k = k' \wedge \langle \forall i : 1 \leq i \leq k : A_i = B_i \rangle . \quad \square$$

Expressions and languages are, of course, different in the same way that names and people are different. (“Winston Churchill” is the name of a famous Englishman. The name consists of a forename and a surname, whilst the person has a mother and father, etc.) Definition 3 has been formulated in a way that makes the difference clear. Henceforth however, we are not so precise and we leave it to the reader to determine whether we are referring to the syntactic form of an expression or to the language that is denoted by the expression. So, for example, a less precise formulation of the first clause of definition 3 is

“the expression  $U \cup V$  is *unambiguous* if  $U \cap V = \emptyset$ ”.

We trust that the reader will have no difficulty in understanding what is meant.

An example of unambiguity is the expression  $\{\varepsilon\} \cup (pre^+.S)T$  in definition 1. Obviously  $\{\varepsilon\} \cap (pre^+.S)T = \emptyset$  because  $\{\varepsilon\}$  is the set of words of length zero whilst  $(pre^+.S)T$  contains only words of length at least one. So the “ $\cup$ ” operator is unambiguous. Also obvious on length considerations is that the (implicit) concatenation operator in the expression  $(pre^+.S)T$  is unambiguous. In general, an expression denoting the concatenation of two languages of which one is a subset of  $T^k$  for some  $k$  (i.e. all the words in the language have the same length) is unambiguous. Deterministic finite-state machines also exemplify the use of unambiguous expressions in order to define a language. A deterministic finite-state machine corresponds to a system of equations in languages; the right sides of the equations are disjoint unions of expressions of the form  $\varepsilon$  or  $aU$  (where  $U$  denotes the language recognised by some state of the machine).

The following lemma is the key to evaluating probabilities and expected values in the context of first-past-the-post games. Note how the equations for  $e_p$  resemble the equations for calculating derivatives.

**Lemma 1.** If  $U \cup V$  is an unambiguous expression,

$$h_p.(U \cup V) = h_p.U + h_p.V , \text{ and}$$

$$e_p.(U \cup V) = e_p.U + e_p.V .$$

If  $UV$  is an unambiguous expression,

$$h_p.UV = h_p.U \times h_p.V , \text{ and}$$

$$e_p.UV = h_p.U \times e_p.V + e_p.U \times h_p.V .$$

*Proof.* Straightforward manipulation of quantifier expressions. □

We now consider the consequences of the function  $p$  being a probability distribution. Recall that we use  $S$  to denote a first-past-the-post game. Because it plays an important role in what follows, we use  $N$  throughout to denote  $pre^+.S$ . (The symbol “ $N$ ” is the one used in [GKP94]; it may be read as a mnemonic for “N”ot complete.) With this notation, the two clauses in definition 1 of a first-past-the-post game become:

- (1)  $N \cap S = \emptyset$  , and
- (2)  $N \cup S = \{\varepsilon\} \cup NT$  .

From (2), it is easy to see that  $h_p.T = 1 \Rightarrow h_p.S = 1$ . See the calculation below.

$$\begin{aligned}
 & h_p.S = 1 \\
 = & \quad \{ \text{heading towards (2) in definition of a game,} \\
 & \quad \text{we add } h_p.N \text{ to both sides} \} \\
 & h_p.N + h_p.S = h_p.N + 1 \\
 = & \quad \{ \text{by definition, } 1 = h_p.\{\varepsilon\}; \text{ assumption: } h_p.T = 1 \} \\
 & h_p.N + h_p.S = h_p.N \times h_p.T + h_p.\{\varepsilon\} \\
 = & \quad \{ \text{expressions } N \cup S \text{ and } NT \cup \{\varepsilon\} \text{ are unambiguous,} \\
 & \quad \text{lemma 1} \} \\
 & h_p.(N \cup S) = h_p.(NT \cup \{\varepsilon\}) \\
 = & \quad \{ \text{definition of a game: (2)} \} \\
 & \text{true} .
 \end{aligned}$$

This suggests that, if  $p$  is a probability distribution on  $T$ ,  $h_p$  is a probability distribution on complete plays. This fact appears to be taken for granted in [GKP94] and [GO81]. (At least, we have been unable to find anything that we would recognise as a proof.) We think it is important to make the theorem explicit and provide a proof. (The proof is not calculational because it links the formal definitions with the informal notion of relative frequencies.)

**Theorem 1.** If  $p$  is a probability distribution on the alphabet  $T$  (i.e.  $p.a$  is the relative frequency of the occurrence of symbol  $a$  when the die is thrown and, thus,  $h_p.T = 1$ ) and throws of the die are independent, the function  $h_p$  is a probability distribution on complete plays of a first-past-the-post game  $S$ . Specifically, for an arbitrary word  $A$  in  $S$ ,  $h_p.A$  is the relative frequency that the word  $A$  is a complete play of the game. Moreover,  $h_p$  is a probability distribution on  $2^S$  (the set of subsets of  $S$ ); if  $U \subseteq S$ , then  $h_p.U$  is the relative frequency with which a word in  $U$  occurs as a complete play.

*Proof.* Suppose  $A \in pre^*.S$ . We prove by induction on the length of  $A$  that  $h_p.A$  is the relative frequency with which the word  $A$  occurs as a prefix of a complete play of the game.

When the length of  $A$  is zero,  $A = \varepsilon$ . The empty word occurs in every play of the game. That is, the relative frequency of  $\varepsilon$  as a prefix of a complete play of the game is 1, which equals  $h_p.\varepsilon$  by definition. This proves the basis.

Now suppose the length of  $A$  is at least one. Suppose  $A = Ba$  for some  $B \in T^*$  and  $a \in T$ . Since  $B \in pre^*.S$ , and the length of  $B$  is less than the length of  $A$ , we may assume inductively that  $h_p.B$  is the relative frequency with which the word  $B$  occurs as a prefix of a complete play of the game. But  $B \in pre^+.S$  and so, by definition 1(b),  $h_p.B$  is the relative frequency with which words of the form  $Bb$ , for some  $b \in T$ , occur as a prefix of a complete play. Since  $p.a$  is the relative frequency that  $a$  occurs, the independence assumption implies that  $h_p.B \times p.a$  is the relative frequency with which  $Ba$  occurs as a prefix of a complete play. But  $h_p.A = h_p.B \times p.a$  by definition. In this way, the induction step is verified.

A corollary of this inductive argument and definition 1(a) is that, when  $A$  is a complete play,  $h_p.A$  is the relative frequency of  $A$  among complete plays. (Because of definition 1(a), a complete play only occurs as a prefix of itself and no other plays.)

By the definition of a probability distribution, it is an immediate corollary that the extension of  $h_p$  to subsets of  $S$  is a probability distribution.  $\square$

Note that  $h_p$  is just a function on arbitrary languages. As shown above, it is a probability distribution on  $S$  and on  $2^S$  whenever  $p$  is a probability distribution on  $T$  but we apply it elsewhere to arbitrary languages. An example of where  $h_p$  is used in this way is the following lemma.

**Lemma 2.** Suppose  $S$  is the set of complete plays in a first-past-the-post game and  $N$  is the set of incomplete plays. Suppose the symbols in  $T$  occur with probability distribution given by  $p$ . Then

$$e_p.S = h_p.N \ .$$

*Proof.* First,

$$\begin{aligned} e_p.S &= h_p.N \\ &= \{ \text{heading towards (2) in definition of a game,} \\ &\quad \text{we add } e_p.N \text{ to both sides} \} \\ e_p.N + e_p.S &= e_p.N + h_p.N \ . \end{aligned}$$

But

$$\begin{aligned} &e_p.N + e_p.S \\ &= \{ \text{expression } N \cup S \text{ is unambiguous, lemma 1} \} \\ &e_p.(N \cup S) \\ &= \{ (2) \} \end{aligned}$$

$$\begin{aligned}
& e_p.(NT \cup \{\varepsilon\}) \\
= & \quad \{ \text{expression } NT \cup \{\varepsilon\} \text{ is unambiguous, lemma 1} \} \\
& h_p.N \times e_p.T + e_p.N \times h_p.T + e_p.\{\varepsilon\} \\
= & \quad \{ \begin{array}{l} p \text{ is a probability distribution on } T, \text{ so } h_p.T = 1; \\ \text{also, for each } A \in T, \#A = 1. \text{ So } e_p.T = 1. \\ \text{By definition, } e_p.\{\varepsilon\} = 0. \end{array} \} \\
& h_p.N + e_p.N \quad .
\end{aligned}$$

The lemma follows by combining the two calculations (using symmetry of addition).  $\square$

*Example 2.* If  $S = \{a, ba, bb\}$  and  $p.a = q$  and  $p.b = r$ , where  $q + r = 1$ , then  $N = \{\varepsilon, b\}$  and  $e_p.S = 1 \times q + 2 \times r \times q + 2 \times r \times r = 1 + r = h_p.N$ .

If  $S = \{b\}^* \{a\}$  and  $p.a = q$  and  $p.b = r$ , then  $N = \{b\}^*$ ; so  $e_p.S = (1 - r)^{-1}$ . (Note how much easier it is to use the lemma than to calculate  $e_p.S$  directly from its definition.)  $\square$

### 3 Prefix-free Languages

A requirement on games is that complete plays are prefix-free languages (definition 1(a)). Any language  $V$  can be reduced to a maximal, prefix-free language by selecting the words that have no proper prefixes in  $V$ . Specifically, if  $V$  is a language, the set  $PF.V$ , called the *prefix-free reduction* of  $V$ , is defined by

$$PF.V = V \cap \neg(VT^+) \quad .$$

The element-wise formulation of  $PF.V$  is that, for all languages  $V$  and all words  $C$ ,

$$C \in PF.V \equiv C \in V \wedge \neg \langle \exists D, E : D \in V \wedge E \in T^+ : DE = C \rangle \quad .$$

That is,  $PF.V$  is the set of words in  $V$  that do not have a proper prefix in  $V$ .

*Example 3.* It is sometimes of interest to determine the expected length of a sequence of observations that culminates in a given “pattern”. Patterns are classified as either *block* or *hidden* [FS09]. Formally, let  $A$  be an arbitrary word over the alphabet  $T$ . Then  $PF.T^* \{A\}$  models the process of observing sequences of letters until the word  $A$  first occurs contiguously (i.e. as a “block” pattern). If  $1 \leq n$  and  $A = a_1 a_2 \dots a_n$ , then  $PF.T^* \{a_1\} T^* \{a_2\} \dots T^* \{a_n\}$  models the process of observing sequences of letters until all the letters of  $A$  occur in order but not necessarily contiguously (i.e. as a “hidden” pattern).

Lemma 6 establishes that  $PF.T^*W$  is a first-past-the-post game for arbitrary non-empty set  $W$ . Thus  $PF.T^* \{A\}$  and  $PF.T^* \{a_1\} T^* \{a_2\} \dots T^* \{a_n\}$  are both first-past-the-post games.

(Of course,  $PF.W$  is not a first-past-the-post game for arbitrary non-empty set  $W$ . A simple counter-example is  $W = \{a\}$  since  $PF.\{a\} = \{a\}$ . When  $T \neq \{a\}$  this is not a first-past-the-post game. See example 1.)  $\square$



The following lemma expresses formally the process of “reducing”  $V$  to  $PF.V$ .

**Lemma 3.** Every word in  $V$  has a unique prefix in  $PF.V$ .

*Proof.* Let  $C$  be a word in  $V$ . Consider a linear search of the prefixes of  $C$ , starting with the empty word and iteratively increasing the length of the prefix, to find a word that is an element of  $V$ . The search will eventually terminate successfully because  $C$  is itself such a word. An invariant of the algorithm is that the current prefix is an element of  $\neg(VT^+)$ . The prefix that is found is thus an element of both  $V$  and  $\neg(VT^+)$ . It is clearly unique because any other prefixes of  $C$  are either not in  $V$  or in  $VT^+$ .  $\square$

Several properties of the function  $PF$  will be used later.

**Lemma 4.**  $PF.V$  is prefix-free. That is, for all  $V$  such that  $V \subseteq T^*$ ,

$$pre^+. (PF.V) \cap PF.V = \emptyset .$$

*Proof.* This is, in fact, a corollary of lemma 3 but is proved directly as follows. We have, for all words  $C$ ,

$$\begin{aligned} & C \in pre^+. (PF.V) \cap PF.V \\ \Rightarrow & \{ \text{definition of } pre^+ \} \\ & \langle \exists E : E \in PF.V : E \in \{C\}T^+ \rangle \wedge C \in PF.V \\ \Rightarrow & \{ PF.V \subseteq V \} \\ & \langle \exists E : E \in PF.V : E \in VT^+ \rangle \\ \Rightarrow & \{ PF.V \subseteq \neg(VT^+) \} \\ & \text{false} . \end{aligned} \quad \square$$

**Remark:** The prefix-free reduction of  $V$  is a maximal prefix-free reduction in the sense that it is *prefix-free* (lemma 4) and it is the *largest* prefix-free subset of  $V$ , i.e. for all languages  $U$ ,

$$(U \subseteq V \equiv U \subseteq PF.V) \Leftrightarrow U \cap pre^+.U = \emptyset .$$

**End of Remark**

**Lemma 5.** For all languages  $V$  and  $U$ , the expression  $(PF.V)U$  is unambiguous. That is, for all languages  $V$  and all words  $C, C', D$  and  $D'$ ,

$$CD = C'D' \wedge C \in PF.V \wedge C' \in PF.V \Rightarrow C = C' \wedge D = D' .$$

*Proof.* We begin with a simple property of words.

$$\begin{aligned} & CD = C'D' \\ \Rightarrow & \{ \text{case analysis on } \#C \text{ and } \#C', \text{ definition of } pre^+ \} \\ & C = C' \vee C \in pre^+.C' \vee C' \in pre^+.C . \end{aligned}$$

We now show that, assuming  $C \in PF.V \wedge C' \in PF.V$ , the second and third disjuncts are false.

$$\begin{aligned}
& C \in pre^+.C' \wedge C' \in PF.V \\
\Rightarrow & \quad \{ \text{definition of } pre^+ \} \\
& C \in pre^+.(PF.V) \\
\Rightarrow & \quad \{ \text{lemma 4} \} \\
& \neg(C \in PF.V) .
\end{aligned}$$

We conclude that

$$C \in pre^+.C' \wedge C \in PF.V \wedge C' \in PF.V \equiv \text{false} .$$

Interchanging the roles of  $C$  and  $C'$ , the third disjunct is also false. The lemma follows straightforwardly.  $\square$

## 4 Block Patterns and Penney-Ante Games

We now specialise the analysis to block patterns and Penney-Ante-type games. In Penney-Ante games, each player chooses a word. A die (with  $|T|$  faces each of which bears one of the elements of  $T$ , but not necessarily fair) is then thrown repeatedly until one of the chosen words occurs as a suffix of the play. The player who made the choice is declared the winner. For example, suppose the alphabet has two symbols  $a$  and  $b$ , one player chooses the word  $a$  and the second player chooses the word  $bb$ . There are just three complete plays of this game: the words  $a$ ,  $ba$  and  $bb$ . The first player wins in the first two cases and the second player wins in the third case. Note that this is a first-past-the-post game — see example 1. Recognition of a block pattern (see example 3) is a special case of a Penney-Ante game with one player.

Consider a set  $W$  of words over an alphabet  $T$ . Note that we do not assume at this stage that  $W$  is finite.

The set  $S$  is defined to be the set of minimal-length words that end in a word in  $W$ . Formally, (in standard regular-language notation)

$$S = T^*W \cap \neg(T^*WT^+) .$$

Equivalently,  $S = PF.T^*W$ .

Returning to the example above, taking  $W$  to be  $\{a, bb\}$  we have:

$$S = \{a, b\}^*\{a, bb\} \cap \neg(\{a, b\}^*\{a, bb\}\{a, b\}^+) = \{a, ba, bb\} .$$

In this very simple example, the set  $S$  is finite; this is not the case in general.

**Lemma 6.** For all  $W$  such that  $W \subseteq T^*$  and  $\emptyset \neq W$ ,  $PF.T^*W$  is a first-past-the-post game.

*Proof.* Let  $S$  denote  $PF.T^*W$  and let  $N$  denote  $pre^+.S$ . Then that  $S$  satisfies 1(a) in the definition of a first-past-the-post game,

$$(3) \quad N \cap S = \emptyset ,$$

is immediate from lemma 4 by instantiating  $V$  to  $T^*W$ .

It remains to verify the property 1(b). Now,

$$\begin{aligned}
pre^*.S &= \{\varepsilon\} \cup (pre^+.S)T \\
&= \{ \quad pre^*.S = S \cup pre^+.S \quad , \quad N = pre^+.S \quad \} \\
S \cup N &= \{\varepsilon\} \cup NT \\
&= \{ \quad T \text{ is the alphabet} \quad \} \\
(S \cup N) \cap T^* &= (\{\varepsilon\} \cup NT) \cap T^* \\
&= \{ \quad T^* = \{\varepsilon\} \cup T^+ \quad \} \\
(S \cup N) \cap (\{\varepsilon\} \cup T^+) &= (\{\varepsilon\} \cup NT) \cap T^* \\
&= \{ \quad \text{distributivity of intersection over union,} \\
&\quad \text{assumption: } \emptyset \neq W. \text{ So } \{\varepsilon\} \subseteq S \cup N \quad \} \\
\{\varepsilon\} \cup ((S \cup N) \cap T^+) &= \{\varepsilon\} \cup (NT \cap T^*) \\
&= \{ \quad NT \subseteq T^+ \subseteq T^*, \\
&\quad \text{cancellation property of languages: } \varepsilon \text{ has length 0} \\
&\quad \text{and words in } T^+ \text{ have length at least 1} \quad \} \\
(S \cup N) \cap T^+ &= NT \\
&= \{ \quad \text{definition of set concatenation and equality} \quad \} \\
\langle \forall B, a : B \in T^* \wedge a \in T : Ba \in S \cup N \equiv B \in N \rangle &.
\end{aligned}$$

Now, for all  $B \in T^*$  and  $a \in T$ , we have

$$\begin{aligned}
&Ba \in S \cup N \\
\Rightarrow &\{ \quad \text{definition of } pre \quad \} \\
&B \in pre.(S \cup N) \\
&= \{ \quad S \cup N = pre^*.S \quad \} \\
&B \in pre^+.S \\
&= \{ \quad N = pre^+.S \quad \} \\
&B \in N .
\end{aligned}$$

For the opposite implication, choose an arbitrary word  $C$  in  $W$ . Then, for all  $B \in T^*$  and  $a \in T$ , we have

$$\begin{aligned}
&B \in N \\
&= \{ \quad C \in W \quad \}
\end{aligned}$$

$$\begin{aligned}
& B \in N \wedge BaC \in T^*W \\
\Rightarrow & \{ \text{lemma 3, definition of } S \} \\
& B \in N \wedge \langle \exists k : 0 \leq k \leq \#(BaC) : pre^k.(BaC) \in S \rangle \\
\Rightarrow & \{ pre^*.B \cap S \\
& \subseteq \{ \text{assume: } B \in N \} \\
& pre^*.N \cap S \\
& = \{ pre^*.N = pre^*(pre^+.S) = pre^+.S = N \} \\
& N \cap S \\
& = \{ (3) \} \\
& \emptyset . \\
& \text{That is, assuming } B \in N, \\
& \langle \forall k : \#(aC) \leq k \leq \#(BaC) : \neg(pre^k.(BaC) \in S) \rangle \} \\
& \langle \exists k : 0 \leq k \leq \#C : pre^k.(BaC) \in S \rangle \\
\Rightarrow & \{ \text{range splitting on } k = \#C, \text{ definition of } N \} \\
& Ba \in S \vee Ba \in N \\
= & \{ \text{definition of set union} \} \\
& Ba \in S \cup N . \quad \square
\end{aligned}$$

#### 4.1 Equations in Languages

In this section, we show how to construct from a given language  $W$  a (non-linear) system of simultaneous equations in languages. The system has one equation for each word in  $W$  (which is not necessarily finite); as we show in section 5, these equations together with the equation 1(b) uniquely characterise  $PF.T^*W$ . Although  $W$  need not be finite, we do assume that it is “reduced”, as defined below.

The set  $W$  is said to be *reduced* if, for all words  $A$  and  $B$  in  $W$ ,  $A$  is a subword<sup>1</sup> of  $B$  equivaless  $A$  equals  $B$ . The assumption that  $W$  is reduced is sensible because without it the game would be either unfair or ill-defined — if  $A$  is a proper suffix of  $B$ , the winner of complete play  $B$  is not well-defined, and if  $A$  is a proper subword of  $B$  and not a proper suffix, the player who chooses  $B$  can never win. For example the set  $\{a, ba, bb\}$  in example 2 is not reduced. (If the complete play is  $ba$ , it is not clear whether the winner is the player choosing  $a$  or the player who chooses  $ba$ .) The need for the assumption also appears formally in our calculations.

If  $A$  is a word in  $W$ ,  $S_A$  is defined by

$$S_A = T^*\{A\} \cap \neg(T^*WT^+) .$$

<sup>1</sup>  $A$  is a subword of  $B$  equivaless there are words  $C$  and  $D$  such that  $B = CAD$ .

Note that  $S = \langle \cup A : A \in W : S_A \rangle$ . The language  $S_A$  is the set of complete plays that end in the word  $A$ .

As in lemma 6, the set  $N$  is defined to be the set of all proper prefixes of  $S$  :

$$N = pre^+.S .$$

(It is straightforward to show that  $N = \neg(T^*WT^*)$ . That is,  $N$  is the set of words of which no word in  $W$  is a subword. This is the definition of  $N$  used by Guibas and Odlyzko [GO81].)

The crucial properties of  $S$  and  $N$  are as follows. If  $W$  is reduced then, for all  $A \in W$ ,

$$(4) \quad N\{A\} = \langle \cup B : B \in W : S_B(B \sqsupseteq A) \rangle$$

where

$$(5) \quad B \sqsupseteq A = \{E, F : \#E < \#A \wedge \#F < \#B \wedge BE = FA : E\} .$$

We pronounce  $B \sqsupseteq A$  as  $B$  match  $A$ . Note that, in spite of the symbol by which it is denoted, the match operator is not symmetric. See example 4 below for instances of the match operator and equations (4) and (5).

For the proof of (4), we first note that

$$\begin{aligned} N\{A\} &= \langle \cup B : B \in W : S_B(B \sqsupseteq A) \rangle \\ &\equiv \langle \forall C :: C \in N \equiv \langle \exists B : B \in W : CA \in S_B(B \sqsupseteq A) \rangle \rangle . \end{aligned}$$

(This is a simple application of the definition of equality of sets, set concatenation and set union.)

Now, for all words  $C$  and all words  $A$  in  $W$ ,

$$\begin{aligned} &\langle \exists B : B \in W : CA \in S_B(B \sqsupseteq A) \rangle \\ = &\quad \{ \text{definition of } B \sqsupseteq A \} \\ &\langle \exists B, E, F : B \in W \wedge \#E < \#A \wedge \#F < \#B \wedge BE = FA : CA \in S_B\{E\} \rangle \\ = &\quad \{ \text{word calculus, } \#(XY) = \#X + \#Y \} \\ &\langle \exists B, D, E, F : B \in W \wedge 1 \leq \#D \leq \#B \wedge A = DE \wedge B = FD : CD \in S_B \rangle \\ = &\quad \{ S_B \subseteq T^*\{B\} \} \\ &\langle \exists B, D, E : B \in W \wedge 1 \leq \#D \leq \#B \wedge A = DE : CD \in S_B \rangle \\ = &\quad \{ \#B < \#D \wedge A = DE \wedge CD \in S_B \\ &\quad \Rightarrow \{ S_B \subseteq T^*\{B\} \} \\ &\quad B \text{ is a proper subword of } A \\ &\quad \Rightarrow \{ W \text{ is reduced, } A \in W \text{ and } B \in W \} \\ &\quad \text{false} \} \\ &\langle \exists B, D, E : B \in W \wedge 1 \leq \#D \wedge A = DE : CD \in S_B \rangle \end{aligned}$$

$$\begin{aligned}
&= \{ S = \langle \cup B : B \in W : S_B \rangle \} \\
&\quad \langle \exists D, E : 1 \leq \#D \wedge DE = A : CD \in S \rangle \\
&= \{ \begin{array}{l} (\Rightarrow) \text{ definition of } N, \\ (\Leftarrow) A \in W, \text{ so } CA \in T^*W; \text{ lemma 3} \end{array} \} \\
&\quad C \in N \ .
\end{aligned}$$

This completes the proof of (4).

*Example 4.* Suppose the alphabet has two symbols  $h$  and  $t$ . Suppose the set  $W$  has three elements  $hh$ ,  $ht$  and  $th$ . The set  $S$  is  $\{t\}^*\{hh, ht, th\}$  and the sets  $S_{hh}$ ,  $S_{th}$  and  $S_{ht}$  are, respectively,  $\{t\}^*\{hh\}$ ,  $\{t\}^*\{th\}$  and  $\{t\}^*\{ht\}$ ; the set  $N$  is  $\{\varepsilon\} \cup \{t\}^*\{h, t\}$ .

The following table shows  $B \sqsubseteq A$  for each of the 9 combinations of  $B$  and  $A$ . (Rows are indexed by  $B$  and columns by  $A$ .)

$\sqsubseteq$	$hh$	$ht$	$th$
$hh$	$\{\varepsilon, h\}$	$\{t\}$	$\emptyset$
$ht$	$\emptyset$	$\{\varepsilon\}$	$\{h\}$
$th$	$\{h\}$	$\{t\}$	$\{\varepsilon\}$

The appropriate instances of (4) are thus as follows:

$$\begin{aligned}
N\{hh\} &= S_{hh}\{\varepsilon, h\} \cup S_{th}\{h\} \\
N\{ht\} &= S_{hh}\{t\} \cup S_{ht} \cup S_{th}\{t\} \\
N\{th\} &= S_{ht}\{h\} \cup S_{th}
\end{aligned}$$

(Some simplification has been applied to these equations. So, for example, in the first equation the term  $S_{ht}\emptyset$  has been omitted and, in the second equation,  $S_{ht}\{\varepsilon\}$  has been simplified to  $S_{ht}$ .)

These equations are complemented by the equations:

$$\begin{aligned}
N \cup S &= \{\varepsilon\} \cup N\{h, t\} \\
S &= S_{hh} \cup S_{ht} \cup S_{th}
\end{aligned}$$

The combination of the two sets of equations is the basis for calculating the probabilities of winning a game with three players who each choose the three words  $hh$ ,  $ht$  and  $th$  as the eventual outcome of the game, as we discuss in the next section.  $\square$

## 4.2 Solov'ev's Equation and Conway's Equation

Suppose we are given a probability distribution  $p$  on the elements of the alphabet  $T$ . Suppose  $W$  is a language and  $S$  equals  $PF.T^*W$ . Then, for each word  $A$  in  $W$ ,  $h_p.S_A$  is the relative frequency that a word ending in  $A$  is a complete play of the game (theorem 1). We show how to use (4) to evaluate  $h_p.S_A$  for each  $A$ . In the case that  $W$  has one element, this gives Solov'ev's equation for the expected

length of a sequence of observations culminating in (the “block pattern”)  $A$ ; see theorem 2. In the case that  $W$  has two elements, this gives Conway’s formula for the probability that each person wins in a two-person Penney-Ante game; see theorem 3.

**Lemma 7.** Suppose  $V$  is a function from words in  $W$  to languages. Suppose  $W$  is reduced and finite. Then  $\langle \cup B : B \in W : S_B V_B \rangle$  is unambiguous.

*Proof.* By lemma 5, each term  $S_B V_B$  is unambiguous. Also, for all words  $D, D', E$  and  $E'$ , and all words  $B$  and  $C$  in  $W$ ,

$$\begin{aligned} DE &= D'E' \wedge D \in S_B \wedge D' \in S_C \\ \Rightarrow \quad & \{ \quad S_B \cup S_C \subseteq PF.T^*W, \text{ lemma 5} \quad \} \\ DE &= D'E' \wedge D \in S_B \wedge D' \in S_C \wedge D = D' \\ \Rightarrow \quad & \{ \quad W \text{ is reduced, } S_B \subseteq T^*\{B\}, S_C \subseteq T^*\{C\} \quad \} \\ D &= D' \wedge E = E' \wedge B = C \quad . \quad \square \end{aligned}$$

**Corollary 1.** For all  $A$  in  $W$ ,

$$h_p.N \times h_p.\{A\} = \langle \Sigma B : B \in W : h_p.S_B \times h_p.(B \boxtimes A) \rangle \quad .$$

Also, for all  $A$  and  $B$  in  $W$ ,

$$h_p.(B \boxtimes A) = \langle \Sigma E, F : \#E < \#A \wedge \#F < \#B \wedge BE = FA : h_p.E \rangle \quad .$$

*Proof.* The expression  $N\{A\}$  is obviously unambiguous. So, by lemma 1,  $h_p.N\{A\}$  is the product of  $h_p.N$  and  $h_p.\{A\}$ . Applying  $h_p$  to both sides of (4), this gives the left side of the first equation above. The right side is immediate from lemma 1 and lemma 7.

The second equation is immediate from lemma 1. (Obviously the right side of (5) is unambiguous.)  $\square$

We are now in a position to formulate the theorems attributed to Sovol’ev and Conway. In the statement of the theorems, the binary operator “:” is defined on pairs of words by, for all  $C$  and  $D$ ,

$$C : D = \frac{h_p.(C \boxtimes D)}{h_p.D} \quad .$$

This operator generalises the one with the same name in [GKP94]. See theorem 3, below, for further explanation of the generalisation.

Note that  $C : D$  has no interpretation as a probability. Indeed,  $C : C$  is typically greater than 1; it is the expected length of the first occurrence of block pattern  $C$ , as shown in the next theorem.

**Theorem 2 (Sovol’ev’s formula).** Suppose  $S = PF.T^*\{A\}$ . Then

$$e_p.S = A : A \quad .$$

*Proof.* We have:

$$\begin{aligned}
& e_p.S \\
= & \{ \text{lemma 2} \} \\
& h_p.N \\
= & \{ \text{corollary 1 with } W := \{A\} \text{ (using } h_p.S = 1 \\
& \text{and one-point rule to simplify the summation)} \} \\
& \frac{h_p.(A \overline{\cup} A)}{h_p.A} \\
= & \{ \text{definition} \} \\
& A : A \ . \quad \square
\end{aligned}$$

*Example 5.* Suppose the alphabet has two symbols  $h$  and  $t$  (for heads and tails). Suppose  $k$  is a natural number and  $A$  is the word  $h^k t$  and  $B$  is the word  $h^{k+1}$ . Then

$$\begin{aligned}
A \overline{\cup} A &= \{\varepsilon\} \\
B \overline{\cup} B &= \{j : 0 \leq j \leq k : h^j\}
\end{aligned}$$

Suppose further that  $p.h = q$  and  $p.t = r$ , where  $q+r = 1$ . It follows that

$$\begin{aligned}
h_p.(A \overline{\cup} A) &= 1 \\
h_p.(B \overline{\cup} B) &= \frac{1-q^{k+1}}{1-q}
\end{aligned}$$

Since  $h_p.A$  is  $q^k \times r$  and  $h_p.B$  is  $q^{k+1}$ ,

$$A : A = \frac{1}{q^k \times r}$$

and

$$B : B = \frac{1-q^{k+1}}{(1-q) \times q^{k+1}} \ .$$

It follows from theorem 2 that

$$\frac{e_p.(PF.T^*\{B\})}{e_p.(PF.T^*\{A\})} = \frac{1-q^{k+1}}{q} \ .$$

The expected number of coin-tosses before  $h^{k+1}$  is encountered is thus approximately  $\frac{1}{q}$  times greater than the expected number of coin-tosses before  $h^k t$  is encountered. In the words of [GKP94]: “patterns with no self-overlaps occur sooner than overlapping patterns do!”  $\square$

**Theorem 3.** Suppose  $W = \{A, B\}$ . Suppose  $W$  is reduced. Then

$$\frac{h_p.S_A}{h_p.S_B} = \frac{B : B - B : A}{A : A - A : B} \ .$$



*Proof.* Straightforward instantiation of corollary 1. □

**Corollary 2 (Conway's formula).** If  $A$  and  $B$  have equal length, and  $p$  assigns equal values to each element of  $T$  then

$$\frac{h_p \cdot S_A}{h_p \cdot S_B} = \frac{h_p \cdot (B \overline{\cap} B) - h_p \cdot (B \overline{\cap} A)}{h_p \cdot (A \overline{\cap} A) - h_p \cdot (A \overline{\cap} B)} .$$

(The latter is equivalent to the formula attributed to John Horton Conway in [GKP94] for the odds of  $A$  winning against  $B$  in a Penney-Ante game where a coin is tossed and the probability of a head or tail occurring is  $\frac{1}{2}$ . In Conway's formula, the notation  $B : A$  is used for  $h_p \cdot (B \overline{\cap} A) \times 2^{\#A - 1}$ . It is not clear from the published literature whether or not Conway derived the general formula given in theorem 3.) □

The examples below test the use of theorem 3 on cases where it is easy to predict the relative frequency of occurrence of words in  $S_A$  and in  $S_B$ .

*Example 6.* Suppose the alphabet has two symbols  $h$  and  $t$  (for heads and tails). Suppose  $k$  is a natural number and  $A$  is the word  $h^k t$  and  $B$  is the word  $h^{k+1}$ . Suppose further that  $p.h = q$  and  $p.t = r$ , where  $q+r = 1$ . A simple argument establishes that the relative frequency of  $A$  compared to  $B$  in a Penney-Ante game is  $\frac{r}{q}$ . We can check that this is predicted by theorem 3 as follows. We first calculate that

$$\begin{aligned} A \overline{\cap} A &= \{\varepsilon\} \\ A \overline{\cap} B &= \emptyset \\ B \overline{\cap} A &= \{j : 0 \leq j < k : h^j t\} . \end{aligned}$$

Then.

$$\begin{aligned} h_p \cdot (A \overline{\cap} B) &= 0 \\ h_p \cdot (B \overline{\cap} A) &= \frac{(q^k - 1) \times r}{q - 1} . \end{aligned}$$

Combining these with the calculations in example 5 and substituting in theorem 3 (top formula), we get, for example,

$$B : A = \frac{(q^k - 1) \times r}{(q - 1) \times (q^k \times r)}$$

Hence, applying theorem 3 (top formula) (and a lot of simplification!), we get

$$\frac{h_p \cdot S_A}{h_p \cdot S_B} = \frac{r}{q}$$

as expected. □

*Example 7.* Suppose the alphabet has two symbols  $a$  and  $b$ . Suppose the set  $W$  has two elements,  $A$  and  $B$ , equal to  $a$  and  $bb$ , respectively. Suppose  $p.a = q$  and  $p.b = r$ , where  $q+r = 1$ . As observed earlier,  $PF.(\{a,b\}^* \{a,bb\}) = \{a,ba,bb\}$ . If  $q$  and  $r$  model the relative frequency of occurrences of  $a$  and  $b$ , respectively,

it is clear that the relative frequency of  $S_A$ , which equals  $\{a,ba\}$ , is  $q+r \times q$  and the relative frequency of  $S_B$ , which equals  $\{bb\}$ , is  $r^2$ . Let us check that this is what is predicted by theorem 3.

We calculate that  $A \overline{\cap} A$  equals  $\{\varepsilon\}$ ,  $B \overline{\cap} B$  equals  $\{\varepsilon,b\}$  and both  $A \overline{\cap} B$  and  $B \overline{\cap} A$  equal the empty set. The  $h_p$  values are now easily calculated. Applying theorem 3, we get

$$\frac{h_p \cdot S_A}{h_p \cdot S_B} = \frac{(1+r) \times q - 0}{1 \times r^2 - 0}$$

which simplifies to  $\frac{(1+r) \times q}{r^2}$ . Since  $q+r=1$  and  $h_p \cdot S_A + h_p \cdot S_B = 1$ , it follows that  $h_p \cdot S_A$  equals  $1-r^2$  and  $h_p \cdot S_B$  equals  $r^2$ .  $\square$

The next example is of a game with an infinite number of players.

*Example 8.* Suppose the alphabet has three symbols  $a$ ,  $b$  and  $c$ . Suppose  $W = \{a\}\{b\}^*\{c\}$ . (So each word in  $W$  is of the form  $ab^k c$  for some  $k$ ,  $0 \leq k$ . Note that  $W$  is not finite but it is reduced.) It is easy to verify that  $ab^k c \overline{\cap} ab^k c = \{\varepsilon\}$  and, when  $j \neq k$ ,  $ab^j c \overline{\cap} ab^k c = \emptyset$ . Thus:

$$\begin{aligned} N\{ab^k c\} &= S_{ab^k c} \\ N \cup S &= \{\varepsilon\} \cup N\{a,b,c\} \\ S &= \langle \cup k : 0 \leq k : S_{ab^k c} \rangle \end{aligned}$$

It is immediate from these equations that  $S = N\{a\}\{b\}^*\{c\}$ . However, it is difficult to “solve” them in the sense of determining a regular expression defining  $N$ . Indeed, it is not even clear that there is a unique solution for  $N$ ; see section 5.

Suppose now that  $p.a=q$ ,  $p.b=r$  and  $p.c=s$ , where  $q+r+s=1$ . Then, exploiting the above equation for  $S$ , we obtain:

$$\begin{aligned} h_p \cdot N \times q \times r^k \times s &= h_p \cdot S_{ab^k c} \\ h_p \cdot N + h_p \cdot N \times q \times r^* \times s &= 1 + h_p \cdot N \times (q+r+s) \end{aligned}$$

(where we write  $r^*$  for  $\frac{1}{1-r}$ ). It follows that  $h_p \cdot N = \frac{1-r}{q \times s}$  and  $h_p \cdot S_{ab^k c} = (1-r) \times r^k$ . So the expected length of a game is  $\frac{1-r}{q \times s}$  (which equals  $\frac{1}{q} + \frac{1}{s}$ ) and the probability that the recognised pattern is  $ab^k c$  is  $(1-r) \times r^k$ .  $\square$

## 5 Uniqueness

In the case that a Penney-Ante game has just two players, theorem 3 together with the equation  $h_p \cdot S_A + h_p \cdot S_B = 1$  enables one to calculate both  $h_p \cdot S_A$  and  $h_p \cdot S_B$ . In other words, it is possible to determine the probability that each of the players wins. This raises the question whether or not the system of equations (4) together with the equations

- (6)  $N \cup S = \{\varepsilon\} \cup NT$
- (7)  $S = \langle \cup A : A \in W : S_A \rangle$

(cf. definition 1(b)) viewed as equations in the unknowns  $N$ ,  $S$  and  $S_A$  (for each  $A \in W$ ), has a unique solution independently of the size of  $W$ .

The answer is no. A very simple example demonstrates this fact. Suppose  $T = \{a\} = W$ . Then, since  $a \sqsupseteq a = \{\varepsilon\}$ , we get just two equations (equation (7) is trivial):

$$\begin{aligned} N \cup S &= \{\varepsilon\} \cup N\{a\} \\ N\{a\} &= S \end{aligned}$$

As is easily checked, one solution to these equations is  $N = \{\varepsilon\}$  and  $S = \{a\}$ . (This is the desired solution.) A second solution is  $N = \{a\}^*$  and  $S = \{a\}^+$ .

Note that, although these two equations do not have a unique solution, we can use them to determine  $h_p.N$  and  $h_p.S$ . Specifically, since inevitably  $h_p.a = 1$ , we get the equations:

$$\begin{aligned} h_p.N + h_p.S &= 1 + h_p.N \\ h_p.N &= h_p.S \end{aligned}$$

Unsurprisingly, the expected length of a complete play is 1. (Apply lemma 2.) Note, however, that  $h_p.\{a\}^*$  is undefined. (Recall that  $\{a\}^*$  is a solution for  $N$ .)

It was a surprise to us that the equations in languages do not have a unique solution since Guibas and Odlyzko [GO81] claim that the derived equations in generating functions do have unique solutions. Their argument is based on the fact that, when  $\{A, B\}$  is reduced,  $\varepsilon \in A \sqsupseteq B \equiv A = B$  for all words  $A$  and  $B$ . We have as yet been unable to use this fact to show that the equations in  $h_p$  values have unique solutions and the question remains open. (Whether or not the equations have unique solutions when we add the equation  $N \cap S = \emptyset$  is irrelevant since this property is not reflected in the generating functions.)

## 6 Conclusion

The purpose of this paper has been to fully understand the reasoning behind the derivation of Conway's formula for solving the Penney-Ante game. Our understanding has improved considerably. In [GKP94] equations are formulated for  $h_p.N$  and  $h_p.S_A$  (for each  $A$ ) —albeit using a different notation— and it is claimed that  $h_p.S_A$  is the probability that the event  $A$  occurs. However, this claim does not appear to be properly justified, as evidenced by the fact that no claim is made about the meaning of  $h_p.N$ . Similarly, we find the arguments given by Guibas and Odlyzko [GO81] somewhat difficult to understand because they involve events that can never occur in a first-past-the-post game: Guibas and Odlyzko appear to ascribe meaning to  $h_p.S_B \times h_p.(B \sqsupseteq A)$  as a probability, whereas, for  $B \neq A$ , the frequency that a word in  $S_B(B \sqsupseteq A)$  ever occurs is 0 — the game would be terminated before such an event occurs. Here we have made clear that  $h_p$  is a probability distribution on the sample space  $PF.T^*W$ . In the derivation of theorem 3, the function is also applied to languages not in this sample space, in which case it is typically not a probability distribution.

We make no use whatsoever of generating functions. Generating functions enable one to derive properties related to word length; our derivations show that word length is irrelevant to deriving Conway's formula (and also Solov'ev's formula). Even in the case of calculating the expected length of a complete play of a game, where word length is part of the definition, lemma 2 is all that is needed.

On the other hand, we have been unable to calculate a formula for the standard deviation of the length of complete games (or for other higher-order cumulants). The conclusion would appear to be that the versatility of generating functions is best demonstrated by their use in determining higher-order cumulants.

The fact that we have been unable to establish uniqueness of the system of equations in languages whilst the system of equations in  $h_p$  values does have a unique solution (since the equations in generating functions have a unique solution) requires further investigation.

## Acknowledgement

I am very grateful to the anonymous referees all five of whom gave very detailed and supportive comments on the submitted paper, including correcting some errors. I hope I have done justice to their efforts.

## References

- [Bac86] R.C. Backhouse. *Program Construction and Verification*. Prentice-Hall International, 1986.
- [Bac03] Roland Backhouse. *Program Construction. Calculating Implementations From Specifications*. John Wiley & Sons, Ltd., 2003.
- [FS09] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.
- [GKP94] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete Mathematics : a Foundation for Computer Science*. Addison-Wesley Publishing Company, second edition, 1994.
- [GO81] L.J. Guibas and A.M. Odlyzko. String overlaps, pattern matching and non-transitive games. *Journal of Combinatorial Theory, Series A30*, pages 183–208, 1981.
- [GS93] David Gries and Fred B. Schneider. *A Logical Approach to Discrete Math*. Springer-Verlag, 1993.
- [Pen74] Walter Penney. Problem 95: Penney-Ante. *Journal of Recreational Mathematics*, page 321, 1974.
- [Sol66] A.D. Solov'ev. A combinatorial identity and its application to the problem concerning the first occurrence of a rare event. *Theory of Probability and its Applications*, 11:276–282, 1966.