Penney Ante: Counterintuitive Probabilities in Coin Tossing

R.S. Nickerson
5 Gleason Road
Bedford, MA 01730
r.nickerson@tufts.edu

Abstract

Coin-tossing is generally viewed as the quintessential example of a random process. We focus on some counterintuitive aspects of sequences that coin-tossing produces. Equally-likely sequences of heads and tails of a specified length are not all equally likely to occur *first*. This realization leads naturally to other surprising facts about coin-tossing. However, aspects of the outcomes of coin-tossing that may be counterintuitive when first encountered can be made acceptable to intuition after reflection and analysis.

Introduction

Reasoning about probabilities can be tricky. Some probability problems are notoriously opaque, even occasionally for people well-versed in probability theory. Examples include

- the three-doors or car-or-goat problem ("the Monty Hall problem") [Vos Savant 1990a; 1990b],
- the sibling-gender problem [Bar-Hillel and Falk 1982],
- the condemned-prisoner problem [Gardner 1961, 226–232],
- Bertrand's paradox [Nickerson 2005], and
- the exchange paradox (two-envelope problem) [Nickerson and Falk 2006].

Even mathematical sophistication does not ensure against difficulty with such problems; the celebrated polymath Paul Erdős refused to accept the correct solution to the three-doors problem when he first encountered it [Hoffman 1998, 253–256; Schechter 1998, 107–109]. These problems and others are discussed in Bar-Hillel and Falk [1982], Falk [1993], and Nickerson [1996; 2004].

The UMAP Journal 28 (4) (2007) 503–532.

The reflections recorded here were triggered by reading Konold [1995], a delightful article in which the author describes his experience of attempting to convince a student of the correctness of an intuition about probability that turned out to be wrong.

The Problem

The question on which Konold's article focuses is:

> Suppose you were to keep flipping a coin until it landed either HTHHT or HHHHH on five consecutive flips. Which of these two sequences would you predict would occur first?

The student picked HTHHT. Konold tried to convince her, with a series of computer-simulated coin-tossing sessions on which they placed wagers at agreed-upon odds, that the two sequences were equally likely to occur first. The student stuck to her belief that HTHHT is more likely than HHHHH to do so. Eventually, Konold became convinced that she was right—she was winning money even when betting at less than 1:1 odds—and was able to construct a compelling argument to that effect. He tipped his hat to his undergraduate student for effectively becoming the tutor to the "would-be instructor."

An Illuminating Case

I find it very easy to identify with Konold's surprise that a strongly-held—and not easily disabused—belief about probability turned out to be wrong, since I have had a similar experience more than once. His change of mind as a consequence of the tutoring experience demonstrates also how easy it is to become confused about a probabilistic relationship even when one thinks that one sees it clearly. In his explanation, he refers to a problem that he remembers being stumped by some years prior to the teaching-learning episode:

> Which would be the most likely result, HH or HT, if you kept flipping a coin until you got one or the other? I remember at first being surprised on discovering that HT was more likely, but it was not hard to see why. With HH, every time you get a T, you are back to square one: You need to flip 2 H's. But with HT, as soon as you get one H, you are "locked in": A T on the next flip will bring success. If instead you get an H, you are still only one T away from success. (p. 206)

That this claim and the explanation of it are wrong was pointed out in several letters to the editor [Abramson 1996; Gottfried 1996; Ilderton 1996; Stone 1996] and good-naturedly acknowledged by Konold [1996]. In fact, HH and HT are equally likely to occur first if one keeps flipping a coin until one of them occurs. Consider the four possible outcomes of the first two tosses: HH, HT, TH, TT. If either HH or HT (which are equally likely) occurs, the game is over. If TH occurs, either an H or a T (again equally likely events) will end the game. If TT occurs, either H or T (equally likely) will occur. If H occurs on the third toss, then either H or T (equally likely) will end it on the fourth. If T occurs on the third toss, then we are essentially in the same position as following the initial tosses of TT. The general point is that a winning combination cannot occur until an H has occurred, and as soon as an H has occurred (independently of how many Ts have preceded it), the next item will end the game, and that item is equally likely to be an H or a T.

As the letter writers noted, the comparison that would have illustrated the desired point is between HH and TH, because TH is more likely than not to be the first of these two sequences to occur if the coin-tossing is continued until one or the other occurs. Consider again the four possible outcomes of the first two tosses: HH, HT, TH, TT. If either HH or TH (which are equally likely) occurs, the game is over. But if either HT or TT (also equally likely) occurs on the first two tosses, the next toss will either complete the game (if H), or cause it to continue (if T), and in the latter case tossing will continue until an H occurs. The general rule is that once a T has occurred, the combination TH is bound to occur before HH does. HH wins this game only if that is the combination on the first two tosses, which has a probability of 1/4, so this game is three times as likely to end with TH than with HH. (A similar analysis will show that HT is more likely to occur first in a string than is TT.)

With longer target sequences the differences in probabilities can become large. Consider, for example, the sequences HHHHH and THHHH. If either of the two target sequences occurs with the first five tosses, the game is over, but if the first five tosses produce any other sequence, it will be impossible for HHHHH to occur before THHHH. It follows that the probability that HHHHH will occur before THHHH is 1/32, or that the odds favoring THHHH are 31 to 1.

Reflection on how one of two equally-likely sequences of heads and tails can be expected to occur before the other more than half the time in coin-tossing exercises leads one quickly to several surprising facts. My purpose in what follows is to consider some of those facts and to attempt to show that they are not implausible, though they may appear to be so at first blush.

A Graphical Representation

The examples HHHHH and THHHH do not prove that HTHHT is more likely than HHHHH to occur first in a sequence of coin tosses (Konold's example), but they perhaps make is easy to believe that that is the case. Konold [1995, Figure 3] presents a graphical representation of the coin-tossing process that demonstrates in another way the plausibility of the likelihood of HTHHT being first. It shows that no matter how many Hs (fewer than 5) have already occurred in a sequence, the occurrence of a T means the process of creating the HHHHH pattern must start from scratch, whereas the occurrence of an H or a T that disrupts the development of the HTHHT pattern does not necessarily mean starting that sequence again from the beginning. If, for example, HTH is followed by T, yielding HTHT, there still remains HT (the first two items of the target sequence) on which to build.

Diagrams similar to that used by Konold and others (e.g., Andrews [2004], Cargal [2003]) are generally referred to as directed graphs, finite-state diagrams or Markov chains; I will call them *state diagrams*. They show the possible paths that could be taken from a beginning state to an end state of some process. The beginning state is the start of the coin-tossing sequence and the end state is the realization of a target sequence. A state diagram for the triplet HHH, for example, is shown in **Figure 1**.

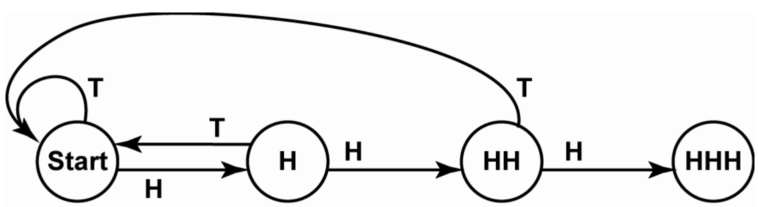


Figure 1. Diagram for HHH.

No matter how many heads (fewer than three) have occurred in sequence, the occurrence of a tail sends the process back to the starting point. The diagram for THH is shown in **Figure 2**.

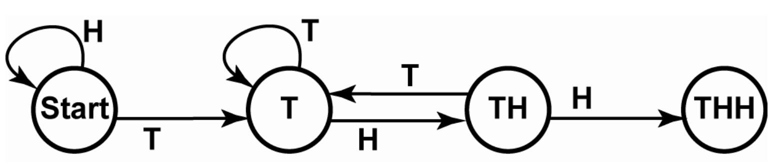


Figure 2. Diagram for THH.

In this case, once a tail has occurred, there is never a return to the starting point.

Expected Waiting Time

The state diagrams for goal states HHH and THH suggest that, on average, it is likely to take longer for HHH to occur than for THH. Each of these graphs has three states at which one of the outcomes will send the process back to a preceding (or the current) state. However, the throwback possibilities are more severe for HHH than for THH: For HHH, one of the two possible outcomes of a toss at every state will move the state back to Start, whereas for THH, at worst an outcome can move the process back to the immediately preceding state.

It is clear from the figures that one can arrive at either goal state by many different paths involving many different numbers of steps. To find the expected number of tosses for any given triplet, one could multiply the length of every unique path to that triplet with its probability of occurrence and take the sum of products over all unique paths. This is impossible, inasmuch as there are infinitely many unique paths to every goal. One can get a sense of relative numbers for specific triplets, however, by considering only paths of moderate length. There are four paths to HHH with five or fewer steps, while there are seven such paths to THH; and the probability that HHH will be reached in not more than five steps is .250 while the probability that THH will be reached in not more than five steps is .375. More generally, THH has more paths leading to it of any given length (except 3) than does HHH, from which it follows that the probability that THH will be reached in n steps, $n > 3$, is greater than the probability that HHH will be reached in the same number of steps.

These considerations lend credence to the idea that expected waiting times can differ across triplets (or across n-tuples, for general n). Techniques for computing expected waiting times (in number of tosses) have been given by Genovese [n.d.], Li [1980], Gerber and Li [1981], Guibas and Odlyzko [1981], and Hombas [1997]. The technique described by Hombas, which I reproduce here, makes use of equations relating expected waiting times for particular sequences conditional on the outcomes of preceding tosses. For example, consider the sequence HTH, the state diagram of which is **Figure 3**.

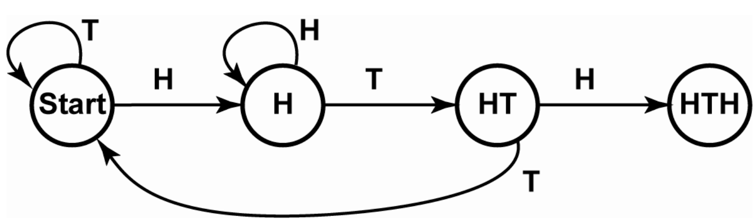


Figure 3. Diagram for HTH.

Letting X represent the number of coin tosses for HTH to occur, Hombas shows that $E(X)$, the expected value of X (expected waiting time for

HTH), can be inferred from a set of linear equations expressing conditional probabilities.

$$E(X) = \tfrac{1}{2}E(X|H_1) + \tfrac{1}{2}E(X|T_1), \tag{1}$$

which is to say that $E(X)$ is the sum of the probabilities of X, conditional on getting H on the first toss and conditional on getting T on the first toss, each multiplied by its probability of occurrence. By reapplying the same reasoning,

$$E(X|H_1) = \tfrac{1}{2}E(X|H_1T_2) + \tfrac{1}{2}E(X|H_1H_2) \tag{2}$$

and

$$E(X|H_1T_2) = \tfrac{1}{2}E(X|H_1T_2H_3) + \tfrac{1}{2}E(X|H_1T_2T_3). \tag{3}$$

Inasmuch as the effect of getting T on the first toss is to stay at Start, thereby increasing $E(X)$ by one, **(1)** can be rewritten as

$$E(X) = \tfrac{1}{2}E(X|H_1) + \tfrac{1}{2}[1 + E(X)]. \tag{4}$$

Similarly, since getting two Hs in a row in effect stalls the process at state H, **(2)** can be rewritten as

$$E(X|H_1) = \tfrac{1}{2}E(X|H_1T_2) + \tfrac{1}{2}[1 + E(X|H_1)]. \tag{5}$$

And, since getting T on the third toss following HT on the first and second tosses sends the process back to Start, **(3)** can be rewritten as[1]

$$E(X|H_1T_2) = \tfrac{1}{2}E(X|H_1T_2H_3) + \tfrac{1}{2}[3 + E(X)]. \tag{6}$$

Solving **(4)** for $E(X|H_1)$ yields

$$E(X|H_1) = E(X) - 1. \tag{7}$$

From **(5)** and **(7)**, we have

$$E(X|H_1T_2) = E(X|H_1) - 1 = E(X) - 2, \tag{8}$$

and from **(6)**,

$$E(X|H_1T_2H_3) = 2E(X|H_1T_2) - 3 - E(X),$$

[1]Making the substitution for the second term on the right can be confusing, especially when solving these equations for longer strings. For X equivalent to THTH, for example, at some point one has the equation $E(X|T_1H_2T_3) = \frac{1}{2}E(X|T_1H_2T_3H_4) + \frac{1}{2}E(X|T_1H_2T_3T_4)$. The correct substitution for $\frac{1}{2}E(X|T_1H_2T_3T_4)$ is $\frac{1}{2}[3 + E(X|T_1)]$. Think of the system as being at the state T_1 but having taken 4 steps (3 more than necessary) to get there. The 3 in the equation represents the 3 wasted steps.

from which, since $E(X|H_1T_2H_3)$ is 3 and, by **(5)**, $E(X|H_1T_2) = E(X) - 2$, we have

$$3 = 2[E(X) - 2] - 3 - E(X),$$

and

$$E(X) = 10. \tag{9}$$

Application of Hombas's technique to all possible triplets shows that their waiting times range from 8 to 14 tosses [Hombas 1997, 131, Table 1c]. HHH and TTT have the longest waiting time, 14 tosses in both cases. HTH and THT have a waiting time of 10, and all other triplets have a waiting time of 8.

Races

Given that different triplets—more generally, different sequences of the same length—can have different waiting times, it seems natural to suspect that different waiting times might account for why one sequence is more likely than another to occur first in a series of tosses. This suspicion might be reinforced by the observation that the waiting time of THH, which is likely to occur before HHH, is only 8, whereas that of HHH is 14. The "race" between HHH and THH may be represented graphically by combining the state diagrams for two sequences in a single diagram, as in **Figure 4**.

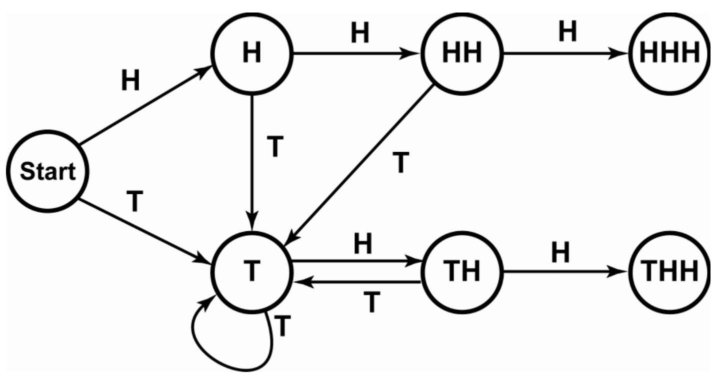


Figure 4. State diagram for the "race" between HHH and THH.

This diagram makes it clear why THH is likely to occur before HHH by showing that once the sequence gets to state T, there is no path that leads to the end state HHH. The state T on the path to THH is an instance of what I call a *clinch state*, a state from which it is impossible to return to a state that is on a path to the alternative end state. A clinch state differs from the conventional absorbing state of a Markov chain, which is a state that once entered is never left.

Applying Hombas's algorithm to HTHHT and HHHHH, the sequences considered by Konold and his student, yields expected waiting times of 44 tosses for the first and 62 for the second. That the sequence that wins the race has the shorter expected waiting time comes as no surprise. What could be more natural than that

- a sequence with a shorter expected waiting time must be likely to occur before a sequence with a longer expected waiting time, and
- two sequences with the same expected waiting time should be equally likely to occur first?

Both of these surmises turn out to be *false*. I will consider the second one first and return to the first one later.

For both HHT and HTT, the expected waiting time is 8; but HHT is expected to occur before HTT 2 times out of 3 on average. The advantage that HHT has over HTT resides in the fact that HH is a clinch state. Once the process gets to that point, it can never get to HTT; it will either get to HHT or recycle at HH. In contrast, even when at its penultimate state on the path to HTT (i.e., at HT), the process can be sent back to a preceding state (H) from which it can eventually get to the opposing end state (HHT).

Another way of representing the race situation is with a standard Markov analysis [Doyle and Snell 1984; Grinstead and Snell 1997, 428–429]. The following description is from Cargal [2003]. The state-to-state transition probabilities shown in **Figure 4** can be represented in matrix form as in **Table 1**.

Table 1. State-to-state transition probabilities for the state diagram of **Figure 4**.

	HHT	HTT	Start	H	HH	HT
HHT	1	0	0	0	0	0
HTT	0	1	0	0	0	0
Start	0	0	.5	.5	0	0
H	0	0	0	0	.5	.5
HH	.5	0	0	0	.5	0
HT	0	.5	0	.5	0	0

A cell entry is the probability of a transition from the state indicated by its row heading to the state indicated by its column heading. So, for example, .5 in the cell (row-column) Start-H indicates that when the process is at Start, the probability of it advancing to H on the next toss is .5. HHT and HTT are *absorbing* states; the game ends when either of these states is attained, so the transition probabilities from either of them to all others is 0.

The matrix, like the corresponding diagram, shows the possibilities at each step in the race to an end state and their probabilities. What one really wants to know, though, is the probability of getting from Start to each of the end states. Cargal [2003] gives a standard procedure for determining these probabilities. Using his notation, we represent the lower left (4-by-2) submatrix by M and the lower right (4-by-4) submatrix by T. The procedure requires that we:

- produce a new matrix, $I - T$, by subtracting T from an identity matrix, I;
- find $(I - T)^{-1}$, the inverse of $I - T$; and
- multiply M by $(I - T)^{-1}$.

The resulting matrix R contains the (multistep) transition probabilities from Start and all intermediate states to each of the end states. For the example being considered, the resulting R is shown in **Table 2**.

Table 2. Solution matrix for the state diagram of **Figure 4**.

	HHT	HTT
Start	.67	.33
H	.67	.33
HH	1	0
HT	.33	.67

This analysis shows that the process is twice as likely to end with HHT as it is to end with HTT. The matrix representation, like the state-diagram, reveals also the advantage that HHT has by virtue of the existence of a clinch state, HH, from which one cannot get to HTT, and no corresponding clinch state for HTT. A clinch state is represented in an R matrix by a row that contains a 1 and a 0 (or when there are $n > 2$ end states, a single 1 and $(n - 1)$ 0s).

The state diagram in **Figure 5** represents the race discussed by Konold. A salient feature is the relatively large number of links to the state HT. Given that HT is on the most direct path to HTHHT, one might take the relatively high probability of getting to HT as suggestive of an advantage of HTHHT over HHHHH in this race. The state-to-state transition matrix for this race is in **Table 3**. Applying the procedure described by Cargal [2003] yields the solution matrix of **Table 4**. We find that HTHHT is nearly twice as likely to win as HHHHH, bearing out what Konold and his student observed in their experiment. This example illustrates also that it is not necessary for a sequence to have a clinch state in order to be a winner; in this case, neither sequence has a clinch state.

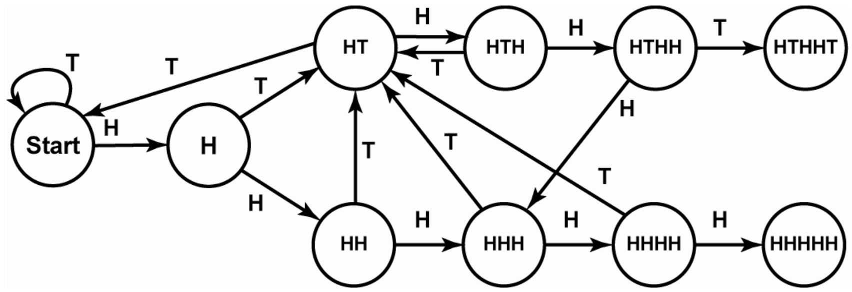


Figure 5. State diagram for the "race" described by Konold, HTHHT vs. HHHHH.

Table 3. Transition matrix for the state diagram of **Figure 5**.

	HTHHT	HHHHH	Start	H	HH	HT	HHH	HTH	HHHH	HTHH
HTHHT	1	0	0	0	0	0	0	0	0	0
HHHHH	0	1	0	0	0	0	0	0	0	0
Start	0	0	.5	.5	0	0	0	0	0	0
H	0	0	0	0	.5	.5	0	0	0	0
HH	0	0	0	0	0	.5	.5	0	0	0
HT	0	0	.5	0	0	0	0	.5	0	0
HHH	0	0	0	0	0	.5	0	0	.5	0
HTH	0	0	0	0	0	.5	0	0	0	.5
HHHH	0	.5	0	0	0	.5	0	0	0	0
HTHH	.5	0	0	0	0	0	.5	0	0	0

Table 4. Solution matrix for the state diagram of **Figure 5**.

	HTHHT	HHHHH
Start	.63	.37
H	.63	.37
HH	.58	.42
HT	.67	.33
HHH	.50	.50
HTH	.71	.29
HHHH	.33	.67
HTHH	.75	.25

Other methods for determining the odds of one specified sequence beating another in a race have been developed, among them one by Conway, described by Gardner [1974]. Li [1980], Grinstead and Snell [1997, 432], and Felix [2006] discuss Conway's algorithm and present some algorithms of their own.

A Counterintuitive Nontransitivity

It should be clear at this point how two sequences of equal length can have different probabilities of occurring first in a series of coin tosses. What may come as a surprise is the fact, first noted by Penney [1969] and described in print by Bostick [1967] (with application to cryptanalysis), that given a specified triplet, one can always find another that has an advantage over it in an extended series of tosses. There is no grand winner—one that will beat all others—in such races. This is a surprising fact; Gardner [1974] claims that most mathematicians cannot believe this when they first hear of it. The situation, generally treated as a game, "Penney Ante," has prompted much discussion [Chrzastowski-Wachtel and Tyszkiewicz 2005; Felix 2006; Graham et al. 1994; Noonan and Zeilberger 2005].

Table 5 shows for every possible pairing of three-item sequences the probability that the one represented by the row (Triplet B) will occur before the one represented by the column (Triplet A). In 20 of the cases, the probability is 1/2; but in all the others, one sequence is more likely than the other to occur first. The table is from Gardner [1974, 123].

Table 5.
Probabilities that sequence B will occur before sequence A in a series of coin tosses (after Gardner [1974, 123]).

	A							
B	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
HHH	*	1/2	2/5	2/5	1/8	5/12	3/10	1/2
HHT	1/2	*	2/3	2/3	1/4	5/8	1/2	7/10
HTH	3/5	1/3	*	1/2	1/2	1/2	3/8	7/12
HTT	3/5	1/3	1/2	*	1/2	1/2	3/4	7/8
THH	7/8	3/4	1/2	1/2	*	1/2	1/3	3/5
THT	7/12	3/8	1/2	1/2	1/2	*	1/3	3/5
TTH	7/10	1/2	5/8	1/4	2/3	2/3	*	1/2
TTT	1/2	3/10	5/12	1/8	2/5	2/5	1/2	*

According to the table, HHT beats HTT (2/3), which beats TTH (3/4), which beats THH (2/3), which beats HHT (3/4). The state diagrams of **Figures 6–9** show the four relevant pairings.

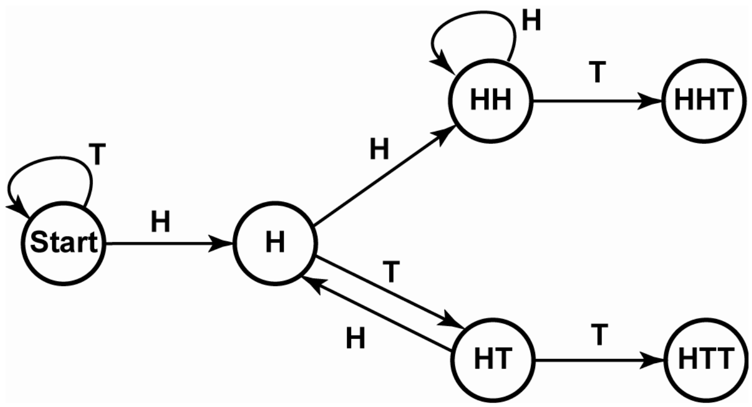


Figure 6. HHT beats HTT.

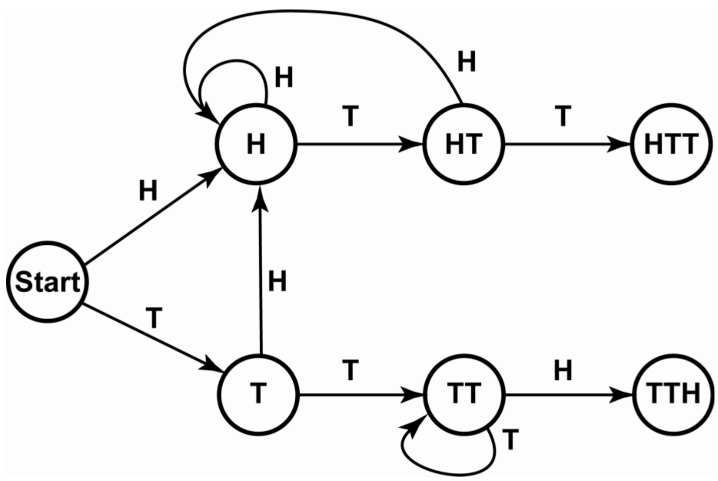


Figure 7. HTT beats TTH.

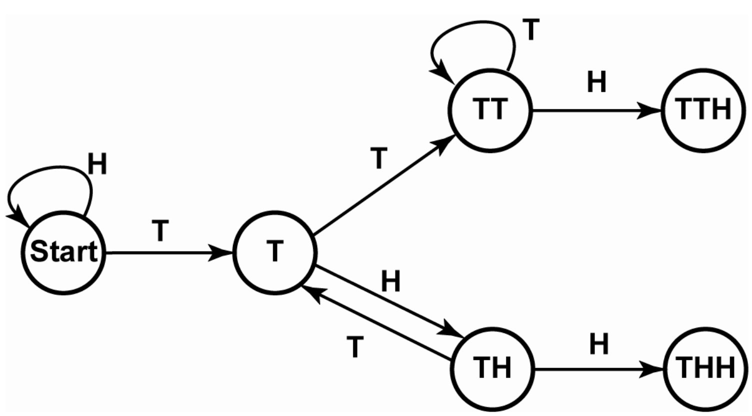


Figure 8. TTH beats THH.

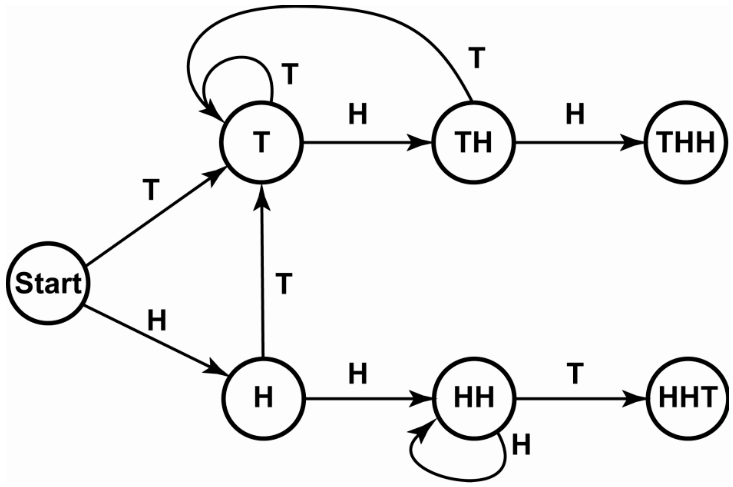
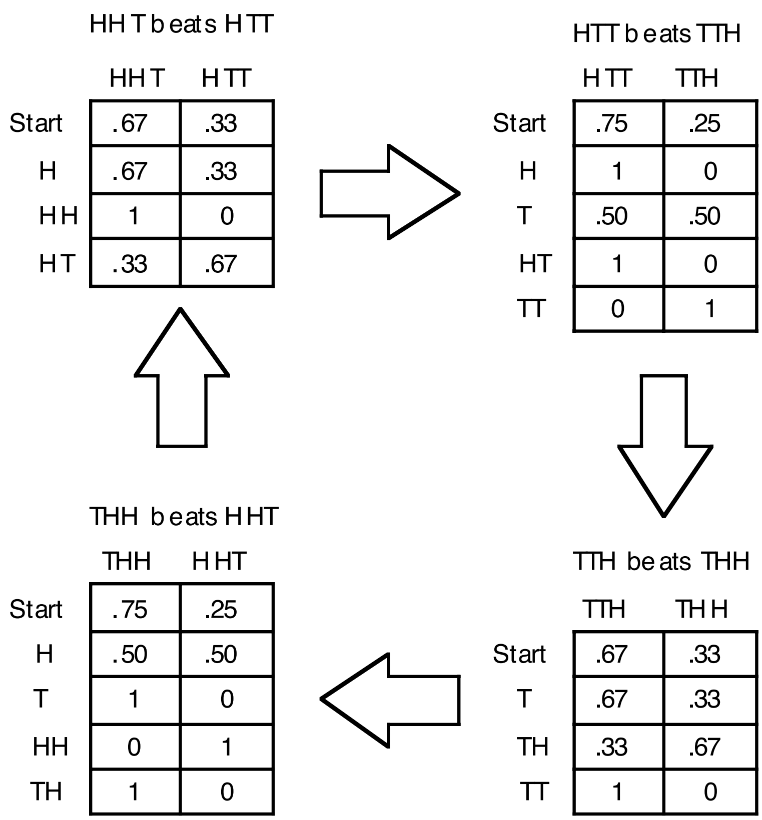


Figure 9. THH beats HHT.

The same cycle may be represented also in terms of R matrices as in **Figure 10**.



HHT beats HTT

	HHT	HTT
Start	.67	.33
H	.67	.33
HH	1	0
HT	.33	.67

HTT beats TTH

	HTT	TTH
Start	.75	.25
H	1	0
T	.50	.50
HT	1	0
TT	0	1

TTH beats THH

	TTH	THH
Start	.67	.33
T	.67	.33
TH	.33	.67
TT	1	0

THH beats HHT

	THH	HHT
Start	.75	.25
H	.50	.50
T	1	0
HH	0	1
TH	1	0

Figure 10. The cycle of nontransitivity of HHT, HTT, TTH, THH, and HHT of **Figures 6–9**.

Each of the races contains one or more clinch states. In HHT beats HTT, there is one clinch state, HH. In HTT beats TTH, there are two relevant clinch states: H for HTT and TT for TTH (HT is also a clinch state, but it is superfluous, inasmuch as HTT is already clinched with the occurrence of H). In all of the analyses shown above, either

- the winning path is the only one that has a clinch state, or
- if both paths have a clinch state, the first clinch state on the winning path has a higher probability of occurring than (and is therefore likely to occur before) the one on the losing path.

The existence of clinch states is specific to the *pairs* that are competing and not to the items comprising a pair. For example, HTT has a clinch state at H when racing against TTH, but it has no clinch state when racing against HHT. This helps make intuitive sense of the possibility of the type of nontransitivity illustrated by the four races just considered.

The nontransitive relationship that we have been considering is not limited to three-item sequences. Indeed, it holds for sequences of any length. Gardner [1974] gives for quadruplets a table comparable to **Table 4**. The nontransitivity among quadruplets is seen in the fact that HTHH beats HHTT (4/7), which beats HTHT (5/9), which beats THTT (9/14), which beats HTTH (7/12), which beats THTH (9/16), which beats HTHH (9/14). In fact it is possible to find intransitivity with as few as three quadruplets. A case in point is the trio HTTH, which beats TTHT (7/12) which beats THTT (3/5) which beats HTTH (7/12) [Chrzastowski-Wachtel and Tyszkiewicz 2005, 150].

A Further Challenge to Intuition

The waiting times of all the triplets in **Figures 6–9** are the same, 8. There is no instance in **Table 5** of a triplet beating one with a shorter waiting time. That one sequence may beat another with the same waiting time is perhaps not much of a strain on credulity, and recognition of the locations of clinch states may suffice to relieve whatever strain there is. But what about the possibility that a sequence would beat another with a shorter waiting time?

There *are* n-tuples, say A and B, for which the expected waiting time for A is shorter than the expected waiting time for B, but B is more likely than A to occur first in a sequence of coin tosses. Gardner [1974] credits Barry Wolk with discovery of this fact and with the identification of the 4-tuples HTHH and THTH as such a pair. Application of Hombas's technique to HTHH and THTH shows them to have expected waiting times of 18 and 20 respectively, but THTH will beat HTHH in a race about 9 times out of 14. Comparison of the state diagrams for those sequences (**Figures 11–12**) makes it easy to see why the expected wait is shorter for HTHH than for THTH.

The salient difference between the possible paths to the end states is that from the penultimate state of HTHH (HTH), the process will either move

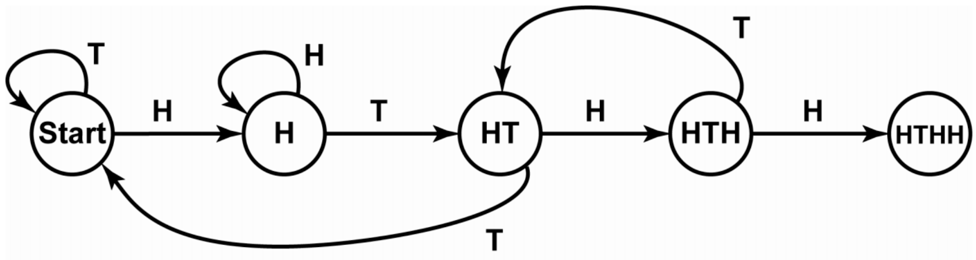


Figure 11. HTHH.

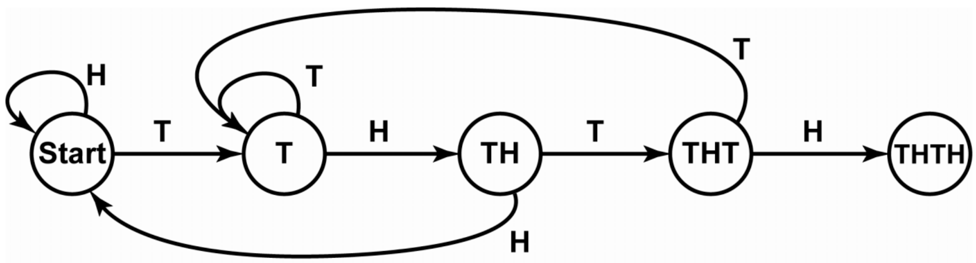


Figure 12. THTH.

to the end state or to the immediately preceding state (HT), whereas from the penultimate state of THTH (THT) the process will either move to the end state or to the state antecedent to the immediately preceding state (T).

The challenge is to make intuitive sense of the fact that the sequence with the longer expected waiting time is more likely to win in a two-way race. The state diagram for this race (**Figure 13**) may help. It shows that the advantage for THTH lies in the fact that from the state penultimate to HTHH, one can get to THTH in only two steps, whereas when in the state penultimate to THTH, one is at least six states away from HTHH. The race between HTHH and THTH is represented in matrix form in **Table 6**. Applying Cargal's algorithm yields the matrix of **Table 7**.

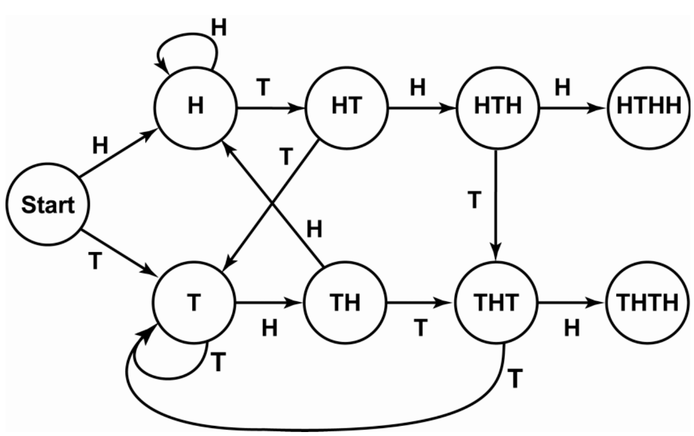


Figure 13. State diagram for the "'race" between HTHH and THTH.

THTH will beat HTHH about 64 times in 100, or about 9 times out of

Table 6. Transition matrix for the state diagram of **Figure 13**.

	HTHH	THTH	Start	H	T	HT	TH	HTH	THT
HTHH	1	0	0	0	0	0	0	0	0
THTH	0	1	0	0	0	0	0	0	0
Start	0	0	0	.5	.5	0	0	0	0
H	0	0	0	.5	0	.5	0	0	0
T	0	0	0	0	.5	0	.5	0	0
HT	0	0	0	0	.5	0	0	.5	0
TH	0	0	0	.5	0	0	0	0	.5
HTH	.5	0	0	0	0	0	0	0	.5
THT	0	.5	0	0	.5	0	0	0	0

Table 7. Solution matrix for the state diagram of **Figure 13**.

	HTHH	THTH
Start	.36	.64
H	.43	.57
T	.29	.71
HT	.43	.57
TH	.29	.71
HTH	.57	.43
THT	.14	.86

14, as already noted. The diagram, and the matrix, illustrate also that it is possible to reach either goal from any non-terminal state—there are no clinch states. It is not necessary for a sequence to have a clinch state to be a winner.

Comparison of the diagrams for the possible paths to the two end states, considered individually, with that representing a race between the two sequences, illustrates why the relationship between expected waiting times does not allow us to predict which sequence is likely to occur first. Another illustration may help make the point.

Imagine a biased coin, say a coin with probability .9 of coming up heads. Toss it until it produces the sequence HT, and then tossing it again until it produces TH. Inspection of the state diagrams for end states HT (**Figure 14**) and TH (**Figure 15**) will convince one that the first case is likely to get quickly to H and to spend considerable time in that state before going to

HT, whereas the second is likely to spin a while on Start before going to T, from where it is likely to proceed quickly to TH, but the expected waiting times for the two sequences are the same.

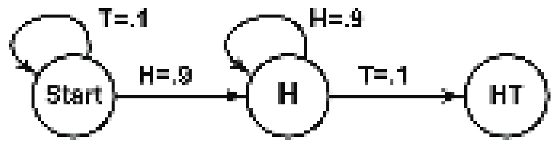


Figure 14. HT.

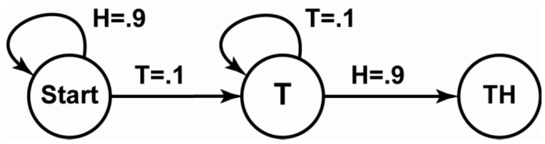


Figure 15. TH.

Neither has an advantage over the other in terms of how long it is likely to take to occur. When HT and TH compete in a race, however, HT will win about 9 times in 10, because both H and T are clinch states, so whether HT or TH wins the race is determined by the outcome of the first toss, which will be H with probability .9 (**Figure 16**).

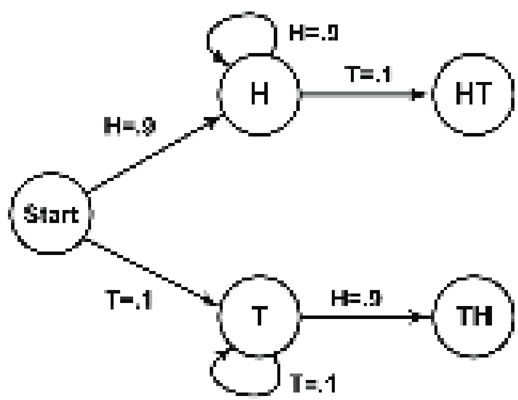


Figure 16. HT vs. TH.

Despite the fact that HT beats TH so decisively, HT will not occur more frequently in a long sequence of coin tosses. Any sequence of tosses can be partitioned into a sequence of runs. In the example, most long runs will be runs of heads, and most runs of tails will be runs only one item long. However, whenever there is a change from a run of heads to a run of tails, the sequence HT occurs; whenever there is a change from a run of tails to a run of heads, the sequence TH occurs; and in any sequence of tosses, the number of transitions from heads to tails must be the same (plus or minus 1) as the number of transitions from tails to heads. So the number of occurrences of HT equals the number of occurrences of TH, plus or minus 1.

This illustration does not demonstrate the possibility of a sequence beating another with a shorter expected waiting time, because the expected waiting times of HT and TH are equal; but it should help to sharpen the distinction between expected waiting time and competitive standing in a race. The following thought experiment should not only sharpen further the distinction, but also show clearly the possibility of a process with a longer expected waiting time beating one with a shorter one. Imagine two processes—A, which requires 1 step two-thirds of the time and 7 steps one-third of the time, and B, which always requires 2 steps. A has an expected waiting time, in steps, of $(2/3)(1) + (1/3)(7) = 3$, whereas B's expected waiting time is 2. So despite the fact that A has a longer expected waiting time than B, it will require less time to finish two times out of three.

To this point, we have considered only races between sequences of the same length. Another jolt to intuition may come from the realization that it is possible to identify sequences of different lengths for which the longer sequence will be more likely to occur before the shorter one [Gardner 1974]. Consider, for example, the sequences HH and THH: HH will win if and only if the first two tosses are both H, which will be the case one-fourth of the time. Similarly, HHH will win a race against THHH if and only if the first three tosses are Hs, which will be the case one-eighth of the time. It should be clear that by racing two sequences that differ in length by one item, the shorter of which is all heads (or all tails) and the longer of which is one tail followed by all heads (or one head followed by all tails), one can make the odds in favor of the longer sequence as large as one wants. It is not necessary that the sequences be as homogeneous as these examples for the longer to have an advantage over the shorter, but these examples make the case.

It is time to emphasize a point. A *race* has connoted the question of which of two specified sequences would occur first when a *single coin* is tossed until one or the other occurs. If we were to define a race as a situation in which two players toss different coins, the winner being the one whose coin produces his or her specified sequence in the lesser number of tosses, then the counter-intuitive relationships that have been described *do not occur*. In this case, the winner in the long run is the player whose specified sequence has the shorter expected waiting time; and if the expected waiting times are the same, the most likely outcome is a tie. Had Konold and his student each tossed his/her own coin, the student still would have won more often than not, because her sequence had the shorter expected waiting time.

Picking Sequences

Several rules have been described for picking an n-item sequence that will beat (in the long run) another n-item sequence that has already been picked [Andrews 2004; Felix 2006; Gardner 1974]. My paraphrase of the one proposed by Andrews [2004] is:

> Make your first item the opposite of the second item in the other sequence (i.e., if the second item in the other sequence is H, make your first item T; if the second item of the other sequence is T, make your first item H); add to your first item the entire other sequence minus the last item in it.

For example, if your opponent in a game of this sort picks HTH, you pick HHT. If you are playing with 5-item sequences and she picks HHTTH, you pick THHTT. **Table 8** shows what Player 2 should pick for each of the possible picks of triplets by Player 1, according to Andrews's rule. The third row of the table gives the odds in favor of Player 2, given the indicated picks, according to **Table 4**.

Table 8.
Player 2's picks to beat Player 1's picks, according to Andrews's rule, and the odds favoring Player 2.

Player 1	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
Player 2	THH	THH	HHT	HHT	TTH	TTH	HTT	HTT
Odds for Player 2	7-1	3-1	2-1	2-1	2-1	2-1	3-1	7-1

Why does this strategy work? One might guess that it works because it ensures that the second sequence to be picked has a higher-probability clinch state than does the first sequence to be picked—that is, a clinch state that occurs earlier than any others in the paths. This is indeed the case for all the pairs in **Table 8**. If one starts with any triplet, and applies Andrews's rule to find a triplet that beats it, applies the rule again to find a triplet that beats that one, and continues in this fashion, one quickly gets into a cycle with the four sequences illustrated in **Figure 10**. Once one gets to any of the triplets, THH, TTH, HTT, or HHT, one is in a nontransitive cycle that could go on indefinitely by successive application of Andrews's rule.

Multisequence Races

To this point, we have considered only races between two specified sequences. The state diagram of **Figure 17** shows a four-way race among the triplets involved in the nontransitive relationships between all pairs of them. The diagram suggests that there is no clear winner when all four triplets compete at once.

This conjecture is borne out by analysis. The transition probabilities in the four-way race are shown in matrix of **Table 9**. Again applying Cargal's algorithm yields the submatrix of **Table 10**, which indicates that the race is a four-way tie.

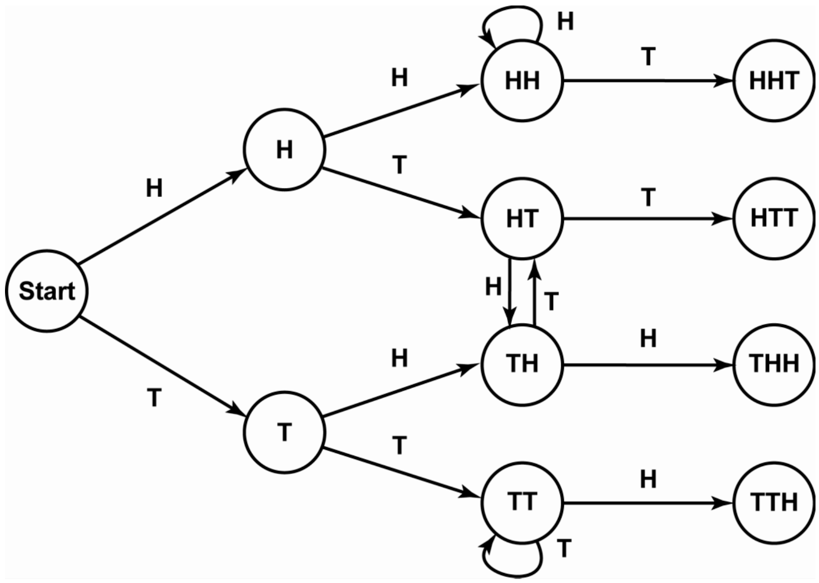


Figure 17. State diagram for a four-way race.

Table 9. Transition probabilities for the state diagram of **Figure 17**.

	HHT	HTT	THH	TTH	Start	H	T	HH	HT	TH	TT
HHT	1	0	0	0	0	0	0	0	0	0	0
HTT	0	1	0	0	0	0	0	0	0	0	0
THH	0	0	1	0	0	0	0	0	0	0	0
TTH	0	0	0	1	0	0	0	0	0	0	0
Start	0	0	0	0	0	.5	.5	0	0	0	0
H	0	0	0	0	0	0	0	.5	.5	0	0
T	0	0	0	0	0	0	0	0	0	.5	.5
HH	.5	0	0	0	0	0	0	.5	0	0	0
HT	0	.5	0	0	0	0	0	0	0	.5	0
TH	0	0	.5	0	0	0	0	0	.5	0	0
TT	0	0	0	.5	0	0	0	0	0	0	.5

Table 10. Solution probabilities for the state diagram of **Figure 17**.

	HHT	HTT	THH	TTH
Start	.25	.25	.25	.25
H	.50	.33	.17	0
T	0	.17	.33	.50
HH	1	0	0	0
HT	0	.67	.33	0
TH	0	.33	. 67	0
TT	0	0	0	1

One might think that HHT and TTH should have an advantage over HTT and THH because HHT and TTH both have a clinch state whereas neither HTT nor THH does. Note, however, that HTT and THH in effect share two states (we might call them *semi-clinch states*) that guarantee that the process will end at one or the other of these end states—if it gets to either HT or TH, it cannot end in either HHT or TTH. And because arriving first at either HT or TH is equally as likely as arriving first at HH or TT, this ensures that the process is as likely to terminate at either HTT or THH as at either HHT or TTH.

Now the challenge to intuition is to reconcile the result of the four-way race with the results of races between pairs of the four same triplets. If all four triplets are equally likely to win a race in which all of them compete, how can one be favored over another when the race is between pairs? The critical thing to see is that a race between two sequences is qualitatively different from a race among four sequences. When the race is among four, it is won when any of the four sequences occurs. When it is between two, it is not won until one of those specific two occurs; if one of the sequences not involved in the race occurs before one of those that are involved does, this does not terminate the race, which proceeds until one of the competing sequences occurs.

Some Data

The foregoing deals with what we should expect, according to probability theory, regarding the occurrences of specific sequences in random series of binary elements, such as would be produced by the tossing of a fair coin. As it happens, a colleague, Susan Butler, and I have the record of 30,000 actual coin (U.S. quarter) tosses (not computer-simulated) that were done for another purpose. Here I report some analyses of the outcomes of this set of tosses, as prompted by the theoretical predictions in the foregoing.

Waiting Times

The mean actual waiting time for each sequence was determined indirectly by counting the number of (nonoverlapping) occurrences of the sequence in the toss outcomes and dividing the resulting number by 30,000. This is tantamount to determining waiting times by starting the count for the first sequence with the first toss in our set and continuing up to and including the toss that completed the first occurrence of the target sequence, and then treating the next toss as the beginning of a new set of tosses, and so on. Suppose the search is for HHH and the first 25 tosses produced

T T H H T H H H H T T H T T T H T H H T T H H H T

The waiting time for the first occurrence of this sequence would be 8 (T T H H T H H H) and the waiting time for the second occurrence would be 16 (H T T H T T T H T H H T T H H H). The resulting waiting times (WT) for the eight triplets are shown in **Table 11**. Expected waiting times (EWT) are included for comparison. The reader will note the correspondence between the actual waiting times of HHT and THH, of HTT and TTH, and of HTH and THT. When I first noticed these correspondences, I immediately suspected a bug in the counting program. However, on reflection, two of them are not surprising; in the case of HHT and THH and that of HTT and THH, the waiting times for the members of each pair are constrained to be identical or nearly so. Consider, for example, HHT and THH. (Bear in mind that searches for the two sequences occur independently—one goes through the set looking for all instances of HHT and then one goes through the same set looking for all instances of THH—so when the two overlap, as in HHTHH, one would be counted in one search and the other in the other.) The reader who may wish to try to construct a sequence in which there are two occurrences of HHT without an intervening THH (or two of THH without an intervening HHT) will easily discover that it cannot be done. Consequently, the number of occurrences of one of these sequences in a series of tosses cannot differ from the number of occurrences of the other sequence by more than 1. The same holds true for the sequences HTT and TTH. The same reasoning does not apply to HTH and THT nor to HHH and TTT; so the exact correspondence between the mean waiting times for HTH and THT is fortuitous.

Table 11.
Actual mean waiting time (WT) and expected waiting time (EWT) for each of the possible triplets.

	Triple							
	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
WT	14.12	8.04	10.17	8.05	8.04	10.17	8.05	13.52
EWT	14.00	8.00	10.00	8.00	8.00	10.00	8.00	14.00

Actual mean waiting times were also determined for other sequences discussed in this paper and the results, along with corresponding expected waiting times are shown in **Table 12**.

Two-way Races between Triplets

All eight triplets were raced against each other in 28 pairwise races, which proceeded as follows. The set of 30,000 coin-toss outcomes was scanned for either member of a specified pair of sequences. As soon as one was found, the race was considered over and that sequence was designated the winner. The next race between the same sequences was started with the toss outcome that immediately followed the toss that terminated the

Table 12.

Actual mean waiting time (WT) and expected waiting time (EWT) for the 5-item sequences of Konold [1995] and the 4-item sequences of Gardner [1974] for which the sequence with the longer expected waiting time beats the one with the shorter expected waiting time.

	Konold's sequences		Gardner's sequences	
	HTHHT	HHHHH	HTHH	THTH
WT	36.35	61.71	18.06	19.99
EWT	38.00	62.00	18.00	20.00

preceding race. The number of races that occurred between any given pair of sequences varied from a low of 4,209 for HTH vs. THT to 5,963 for HHT vs TTH; the mean was 5,080. The mean absolute deviation between actual and expected percentages of wins was approximately 0.2%. (See **Tables A1** and **A2** in the **Appendix**.) The nontransitivity with respect to winnings noted in the theoretical part of this paper was observed with the actual race data: HHT beat HTT (3,730 to 1,839; 67.0%), which beat TTH (3,435 to 1,197; 74.2%), which beat THH (3,726 to 1,840; 66.9%), which beat HHT (3,463 to 1,111; 75.7%).

Two-way Race between Konold's Quintuples

The two five-element sequences considered by Konold, HTHHT and HHHHH, were raced. According to theory, HTHHT should win about 63% of the time. The result was that HTHHT actually won 801 (62.3%) of the 1286 races.

Two-way Race between HTHH and THTH

Recall that the expected waiting times for HTHH and THTH are 18 and 20 respectively, and the mean actual waiting times in our sample were 18.06 and 19.99, and that, according to theory, the sequence with the longer expected waiting time should win about 64 percent of the time. In fact, THTH won 1501 (65.3%) of 2296 races.

Four-way Race among Triplets

The four triplets making up the nontransitive cycle were run in a four-way race. The percentages of wins were: HHT: 24.2%, HTT: 24.5%, THH: 25.1%, and TTH: 26.2%.

Number of Occurrences

Konold [1995] ends his paper with an observation with which many people will find it easy to identify. Commenting on the understanding of a problem that one can develop—or that one may believe one has developed—in the course of thinking about it, he notes that "understanding does not typically arrive suddenly like a newborn and set up permanent residence. More like a teenager, it pops in and out" (p. 209). By way of illustrating the point, Konold recounts his further thinking about the problem described at the beginning of his article.

> After I thought I had come to terms with the flip-until problem, the following dilemma set me back momentarily. Suppose I flipped a coin 1,000 times and wrote down the results in one long string. I could search for occurrences of HHHHH and HTHHT by sliding a "window" along the string that allowed me to see only five characters at a time. If I started at the beginning of the string and advanced the window one character at a time, I could view 996 events of length 5. I am convinced that in this sample the expected number of occurrences of HHHHH, HTHHT, or any other sequence of length 5 is 996 $(1/2)^5$. How can this be reconciled with the fact that, sliding the window along, I expect to encounter the first instance of HTHHT before encountering HHHHH? (p. 209)

The number of occurrences of each of the two 5-element sequences considered by Konold was counted in our set of 30,000 tosses. The results were: HTHHT 926 and HHHHH 944, which is to say that (consistent with Konold's assumption) these sequences occurred with nearly equal frequency, close to the expected frequency of approximately 937. (There are 29,996 overlapping 5-element sequences in a set of 30,000 tosses; with all 32 possible sequences occurring with equal frequency, the expected number of occurrences of any specific sequence is approximately 937.)

That all possible sequences of a specified length are expected to occur with about equal frequency in a long series of tosses may itself pose a challenge to intuition. As both Gottfried [1996] and Ilderton [1996] point out in comments on Konold's [1995] article, when counting sequences with the "sliding window," a sequence of all heads (or all tails) can follow itself immediately, whereas a sequence that is a mix of heads and tails cannot. More generally, the probability of occurrence of a specified sequence is not independent of preceding sequences (as it would be if one counted according to what Konold refers to as the "block method," which considers only successive nonoverlapping n-tuples).

Such constraints on successive recurrences accounts for why sequences with different expected waiting times nevertheless occur approximately the same number of times in a large set of tosses. Recall that the expected waiting time for HHH and TTT is 14, that of HTH and THT is 10, and that of all other triplets is 8, as shown in **Table 11**. In the "sliding window" count of number of occurrences, the somewhat long waiting time of HTH and THT—relative to that of HHT, HTT, THH, and TTH—is compensated for by the fact that successive occurrences of each of these triplets can overlap by a single toss; and the even longer waiting time of HHH and TTT is offset by the fact that successive occurrence of each of these can overlap by two tosses. Genovese [n.d.] shows that the greater overlap a sequence has with itself, the longer its expected waiting time. With longer sequences, the difference between expected waiting times as a function of difference in amount of overlap can be quite large. Genovese illustrates this with a comparison between HHHHHHH, which takes on average 254 tosses to occur, and HTHHTTT, which takes only 128, about half as many.

Noonan and Zeilberger [2005] discuss the relationship between the Penney Ante game and the general problem of finding specified words or word segments in running text, and they describe software available from their Website that can be used to determine the winners of specified sequences in the Penney Ante game. They offer also programs that simulate Penney Ante games and identify the best sequence to pick (countermove) given the sequence selected by an opponent. Software for analyzing Penney Ante is also described by Chrzastowski-Wachtel and Tyszkiewicz [2005].

Concluding Comments

I suspect that the occurrence of sequences of heads and tails in a series of coin tosses is not a topic to which most people give a lot of thought. For writers of textbooks on probability theory, however, coin-tossing is the prototypical example of a random process. The phenomena discussed in this article are illustrative of counterintuitive relationships that can hold among probabilistic variables. In particular, they demonstrate that ***all the following assumptions that one might be tempted to make are wrong***:

- Random events that are equally likely to occur are necessarily equally likely to occur first in a sequence of outputs of the random process that produces them.
- Random events with the same expected waiting time are necessarily equally likely to occur first in a sequence of outputs of the random process that produces them.
- If two random events have different expected waiting times, the event with the shorter expected waiting time is necessarily more likely than the other to occur first in a sequence of outputs of the random process that produces them.
- If $n > 2$ random events are equally likely to occur first in a sequence of outputs of the random process that produces them, any two of them are necessarily equally likely to occur first when raced against each other.

- It is not possible for random events to have a relationship of the sort $A < B < C < D < A$, where $A < B$ means that A is likely to occur before B.

Such surprises are compelling evidence of the wisdom of caution in drawing conclusions about probabilistic relationships even if they appear, at first blush, to be obvious.

Appendix

Table A1.

Observed number of races in which B beat A in dataset of 30,000 coin-toss outcomes.

		A							
		HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
	HHH	*	2124	1783	2124	541	2124	1564	2124
	HHT	2147	*	3334	3730	1111	3730	2989	3730
	HTH	2599	1644	*	2949	2453	2103	2212	2949
B	HTT	3194	1839	2991	*	2938	2519	3435	3726
	THH	3730	3463	2531	2998	*	2985	1840	3186
	THT	2952	2209	2106	2483	2952	*	1654	2598
	TTH	3726	2974	3726	1197	3726	3344	*	2100
	TTT	2216	1669	2216	590	2216	1857	2216	*

Table A2.
Observed percentage of races in which B beat A in data set of 30,000 coin-toss outcomes (top line), with expected percentage according to probability theory (bottom line).

		A							
		HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
	HHH	*	49.7 50.0	40.7 40.0	39.9 40.0	12.7 12.5	41.8 41.7	29.6 30.0	48.9 50.0
	HHT	50.3 50.0	*	67.0 66.7	67.0 66.7	24.3 25.0	62.8 62.5	50.1 50.0	69.1 70.0
	HTH	59.3 60.0	33.0 33.3	*	49.6 50.0	49.2 50.0	50.0 50.0	37.3 37.5	57.1 58.3
B	HTT	60.1 60.0	33.0 33.3	50.4 50.0	*	49.5 50.0	50.4 50.0	74.2 75.0	84.4 87.5
	THH	87.3 87.5	75.7 75.0	50.8 50.0	50.5 50.0	*	50.3 50.0	33.1 33.3	59.0 60.0
	THT	58.2 58.3	37.2 37.5	50.0 50.0	49.6 50.0	49.7 50.0	*	33.1 33.3	58.3 60.0
	TTH	70.4 70.0	49.9 50.0	62.7 62.5	25.8 25.0	66.9 66.7	66.9 66.7	*	48.7 50.0
	TTT	51.1 50.0	30.9 30.0	42.9 41.7	15.6 12.5	41.0 40.0	41.7 40.0	51.3 50.0	*

References

Abramson, Lee R. 1996. Comment on Konold [1995]. *American Statistician* 50: 200.

Andrews, Mark W. 2004. Anyone for a nontransitive paradox? The case of Penney-Ante. Preprint. Available from the author at `m.andrews@ucl.ac.uk` or `ucjtman@ucl.ac.uk` .

Bar-Hillel, M.A. and Ruma Falk. 1982. Some teasers concerning conditional probabilities. *Cognition* 11: 109–122.

Bostick, Charles W. Penney-Ante. 1967. *NSA Technical Journal* 12 (1)(Winter 1967) 29–36. Available from Center for Cryptologic History at `history@nsa.gov` .

Cargal, James M. 2003. Cargal's lectures on algorithms, number theory, probability and other stuff. `http://www.cargalmathbooks.com/` .

Chrząstowski-Wachtel, P., and J. Tyszkiewicz. 2005. A Maple package for conditional event algebras. In *Conditionals, Information, and Inference: International Workshop, WCII 2002, Hagen, Germany, May 13–15, 2002,* *Revised Selected Papers*, edited by G. Kern-Isberner, W. Rödder, and F. Kulmann, 131–151. New York: Springer.

Doyle, Peter G., and J. Laurie Snell. 1984. *Random Walks and Electric Networks.* Washington, DC: Mathematical Association of America.

Falk, Ruma. 1993. *Understanding Probability and Statistics: A Book of Problems.* Wellesley, MA: A.K. Peters.

Felix, D. 2006. Optimal Penney Ante strategy via correlation polynomial identities. *Electronic Journal of Combinatorics* 13: 1–15.

Gardner, Martin 1961. *The Second* Scientific American *Book of Mathematical Puzzles and Diversions.* New York: Simon and Schuster. 1987. Reprint. Chicago, IL: University of Chicago Press. 2005. Reissue on CD-ROM. *Martin Gardner's Mathematical Games.* Washington, DC: Mathematical Association of America.

________ [misspelled as Gardener, Martin]. 1972–73. One to one: A correspondence column. Letter. *Manifold: A Mathematics Magazine* (13): 36; `http://www.jaworski.co.uk/m13/13_oneone.html` .

________. 1974. On the paradoxical situations that arise from nontransitive relations. *Scientific American* 231 (4): 120–125. 1988. Reprinted as Chapter 5 in *Time Travel and Other Mathematical Bewilderments*. New York: W.H. Freeman. 2005. Reissue on CD-ROM. *Martin Gardner's Mathematical Games.* Washington, DC: Mathematical Association of America. 2001. Reprinted as Chapter 23 in *The Colossal Book of Mathematics: Classic Puzzles, Paradoxes, and Problems*. New York: W.W. Norton.

Genovese, C. n.d. Case study #2: Pattern matching. `http://signalplusnoise.com/extras/penneyante.pdf` .

Gerber, H., and S-Y.R. Li. 1981. The occurrence of sequence patterns in repeated experiments and hitting times in a Markov chain. *Stochastic Processes and their Applications* 11: 101–108.

Gottfried, Paul. 1996. Comment on Konold [1995]. *American Statistician* 50: 200.

Graham, Ronald L., Donald E. Knuth, and O. Patashnik. 1994. *Concrete Mathematics*. 2nd ed. Reading, MA: Addison-Wesley.

Grinstead, Charles M., and J. Laurie Snell. 1997. *Introduction to Probability*. 2nd rev. ed. Providence, RI: American Mathematical Society.

Guibas, L.J., and Andrew M. Odlyzko. 1981. String overlaps, pattern matching, and nontransitive games. *Journal of Combinatorial Theory* A30: 183–208.

HEC [École des Hautes Études Commerciales HEC]. 2004. Paradoxe de Walter Penney (1969). MATH HEC T. Annale d'examen, Mathématiques I, Classe Prepa HEC (STG), Concours du Supérieur. `http://`

www.bankexam.fr/annales/1634_HEC_2004_concours.pdf . Variations of this exam question appear at http://80.11.80.31/maths/upsmaths/BULLETIN/bult2004/m04al1e.pdf and elsewhere, and comments on the solution papers are at http://www.escp-eap.net/fr/programmes/master/rapports_epreuves_2004/pdf_ecrit/rapport_option_%20maths_bl.pdf .

Hoffman, P. 1998. *The Man who Loved Only Numbers: The Story of Paul Erdős and the Search for Mathematical Truth.* New York: Hyperion.

Hombas, V.C. 1997. Waiting time and expected waiting time—Paradoxical situations. *American Statistician* 51: 130–133.

Ilderton, Robert B. 1996. Comment on Konold [1995]. *American Statistician* 50: 200.

Konold, Clifford. 1995. Confessions of a coin flipper and would-be instructor. *American Statistician* 49: 203–209.

________. Reply to comments on Konold [1995]. *American Statistician* 50: 200.

Li, Shuo-Yen Robert. 1980. A Martingale approach to the study of occurrence of sequence patterns in repeated experiments. *Annals of Probability* 8: 1171–1176.

Nickerson, R.S. 1996. Ambiguities and unstated assumptions in probabilistic reasoning. *Psychological Bulletin* 120: 410–433.

________. 2004. *Cognition and Chance: The Psychology of Probabilistic Reasoning.* Mahwah, NJ: Lawrence Erlbaum Associates.

________. 2005. Bertrand's chord, Buffon's needle, and the concept of randomness. *Thinking and Reasoning* 11: 67–96.

________, and Ruma Falk. 2006. The exchange paradox: Probabilistic and cognitive analysis of a psychological conundrum. *Thinking and Reasoning* 12: 181–213.

Noonan, John, and Doron Zeilberger. 1999. The Goulden-Jackson cluster method: Extensions, applications and implementations. *Journal of Difference Equations and Their Applications* 5 (1999): 355–377; http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/gj.html . Accompanying Maple packages at http://www.math.rutgers.edu/~zeilberg/gj.html .

Penney, Walter. 1969. Problem 95. Penney-Ante. *Journal of Recreational Mathematics* 2 (4)(October 1969): 241. 1974. Solution by the proposer. 7 (4): 321. 1975. Shortened version of solution. 8 (1): 61–62.

Schechter, B. 1998. *My Brain Is Open: The Mathematical Journeys of Paul Erdős.* New York: Simon and Schuster.

Silverman, David L. 1969. Problem 96. Blind Penney-Ante. *Journal of Recreational Mathematics* 2 (4)(October 1969): 241. 1975. Solution by Ned Shilling. 8 (1): 62–65.

Stark, Dudley. 1995. First occurrence in pairs of long words: A Penney-ante conjecture of Pevzner. *Combinatorics, Probability, and Computing* 4 (1995): 279–285.

________. 1971. Problem 177. A Penney-Ante paradox. *Journal of Recreational Mathematics* 4 (4)(October 1971): 241. 1975–76. Solutions by W.W. Funkenbush and R. Robinson Rowe. 8(2): 147–149.

Stone, Dan C. 1996. Comment on Konold [1995]. *American Statistician* 50: 200.

Vos Savant, Marilyn. 1990a. Ask Marilyn. *Parade Magazine* (9 September 1990): 15.

________. 1990b. Ask Marilyn. *Parade Magazine* (2 December 1990): 25.

Acknowledgments

Alexander Lauritson-Lada and Shivani Sheopory wrote the programs for the data analyses reported in this paper. The manuscript benefited from very helpful comments by Susan Butler, two anonymous reviewers, and the journal editor. The work was supported by National Science Foundation grant 0241739 to Raymond S. Nickerson.

About the Author

Raymond S. Nickerson, retired senior vice-president of Bolt Beranek and Newman Inc. and research professor at Tufts University, is a fellow of the American Association for the Advancement of Science, the American Psychological Association, the Association for Psychological Science, the Human Factors and Ergonomics Society, and the Society of Experimental Psychologists. The founding editor of the *Journal of Experimental Psychology: Applied* and the founding and series editor of *Reviews of Human Factors and Ergonomics*, an annual publication of the Human Factors and Ergonomics Society, he is the author of several books, the latest two of which are *Cognition and Chance: The Psychology of Probabilistic Reasoning* (Erlbaum, 2004) and A*spects of Rationality: Reflections on What It Means To Be Rational and Whether We Are* (Psychology Press, 2007).