346 (Mr.Bean's socks) Mr.Bean is trying to get a matching pair of socks from a drawer containing an inexhaustible supply of red and blue socks. He begins by withdrawing two socks at random. If they match, he is done. Otherwise, he throws away one of them at random, withdraws another sock at random, and repeats. How long will it take him to get a matching pair? Assume that a sock withdrawn from the drawer has 1/2 probability of being each color, and that a sock that is thrown away also has a 1/2 probability of being each color.

After trying the question, scroll down to the solution.

§ Informally, here is Mr.Bean's program.

choose a sock color with the left hand.
choose a sock color with the right hand.
loop

where

loop $\Leftarrow$ **if** sock colors match **then** *ok*
else choose a hand and a sock color for that hand. loop **fi**

Let variables L and R represent the color of socks held in Mr.Bean's left and right hands, and let time variable t count iterations. Formally, the program is

if *rand* 2 **then** $L{:=}\ red$ **else** $L{:=}\ blue$ **fi**.
if *rand* 2 **then** $R{:=}\ red$ **else** $R{:=}\ blue$ **fi**.
loop

where

loop $\Leftarrow$ **if** $L{=}R$ **then** *ok*
else **if** *rand* 2 **then** **if** *rand* 2 **then** $L{:=}\ red$ **else** $L{:=}\ blue$ **fi**
else if *rand* 2 **then** $R{:=}\ red$ **else** $R{=}\ blue$ **fi** **fi**.
$t{:=}\ t{+}1$. *loop* **fi**

All occurrences of *rand* 2 can be replaced by $1/2$. Since *red* and *blue* are the only two values for L and R , the two initialization lines can be simplified as follows:

if $1/2$ **then** $L{:=}\ red$ **else** $L{:=}\ blue$ **fi**.
if $1/2$ **then** $R{:=}\ red$ **else** $R{:=}\ blue$ **fi** — replace **if** and :=

$= \quad 1/2 \times (L'{=}red) \times (R'{=}R) \times (t'{=}t) \ + \ 1/2 \times (L'{=}blue) \times (R'{=}R) \times (t'{=}t).$
$\quad\ \ 1/2 \times (R'{=}red) \times (L'{=}L) \times (t'{=}t) \ + \ 1/2 \times (R'{=}blue) \times (L'{=}L) \times (t'{=}t)$

$= \quad ((L'{=}red) + (L'{=}blue)) \times (R'{=}R) \times (t'{=}t)\ /\ 2.$
$\quad\ \ ((R'{=}red) + (R'{=}blue)) \times (L'{=}L) \times (t'{=}t)\ /2$

for either value of L' the sum is 1 , and similarly for R'

$= \quad (R'{=}R) \times (t'{=}t)\ /\ 2.\ \ (L'{=}L) \times (t'{=}t)\ /\ 2$ — replace . and use one-point law

$= \quad (t'{=}t)/4$

Similarly the loop body can be simplified:

if $1/2$ **then if** $1/2$ **then** $L{:=}\ red$ **else** $L{:=}\ blue$ **fi**
else if $1/2$ **then** $R{:=}\ red$ **else** $R{=}\ blue$ **fi fi**

$= \quad ((L'{=}L) + (R'{=}R)) \times (t'{=}t)\ /\ 4$

The program is now

$(t'{=}t)/4$. *loop*

where

loop $\Leftarrow$ **if** $L{=}R$ **then** *ok* **else** $((L'{=}L) + (R'{=}R)) \times (t'{=}t)\ /\ 4.\ \ t{:=}\ t{+}1.$ *loop* **fi**

After three failed attempts to define *loop* I propose

loop $=$ **if** $L{=}R$ **then** *ok* **else** $(L'{=}R') \times (t'{>}t) \times 2^{t-t'}$ **fi**

Here's the proof of the refinement, starting with the right side.

if $L{=}R$ **then** *ok* **else** $((L'{=}L) + (R'{=}R)) \times (t'{=}t)\ /\ 4.\ \ t{:=}\ t{+}1.$ *loop* **fi** — expand *loop*

$=$ **if** $L{=}R$ **then** *ok* **else** $((L'{=}L) + (R'{=}R)) \times (t'{=}t)\ /\ 4.\ \ t{:=}\ t{+}1.$
if $L{=}R$ **then** *ok* **else** $(L'{=}R') \times (t'{>}t) \times 2^{t-t'}$ **fi fi** — expand last *ok*

$=$ **if** $L{=}R$ **then** *ok*
else $((L'{=}L) + (R'{=}R)) \times (t'{=}t)\ /\ 4.\ \ t{:=}\ t{+}1.$
if $L{=}R$ **then** $(L'{=}L) \times (R'{=}R) \times (t'{=}t)$ **else** $(L'{=}R') \times (t'{>}t) \times 2^{t-t'}$ **fi fi** — substitution

$=$ **if** $L{=}R$ **then** *ok*
else $((L'{=}L) + (R'{=}R)) \times (t'{=}t)\ /\ 4.$
if $L{=}R$ **then** $(L'{=}L) \times (R'{=}R) \times (t'{=}t{+}1)$ **else** $(L'{=}R') \times (t'{>}t{+}1) \times 2^{t+1-t'}$ **fi fi**

sequential composition

$=$ **if** $L{=}R$ **then** *ok*
else $\Sigma L'', R'', t''\cdot((L''{=}L) + (R''{=}R)) \times (t''{=}t)\ /\ 4\ \times$
if $L''{=}R''$ **then** $(L'{=}L'') \times (R'{=}R'') \times (t'{=}t''{+}1)$ **else** $(L'{=}R') \times (t'{>}t''{+}1) \times 2^{t''+1-t'}$ **fi fi**

expand **if**

$$
\begin{aligned}
=\;& \textbf{if } L{=}R \textbf{ then } ok \\
& \textbf{else } \Sigma L'', R'', t''\cdot((L''{=}L) + (R''{=}R)) \times (t''{=}t) / 4 \times \\
& \qquad\qquad (\ (L''{=}R'') \times (L'{=}L'') \times (R'{=}R'') \times (t'{=}t''{+}1) \\
& \qquad\qquad + (L''{\ne}R'') \times (L'{=}R') \times (t'{>}t''{+}1) \times 2^{t''+1-t'}) \textbf{ fi}
\end{aligned}
$$

distribute

$$
\begin{aligned}
=\;& \textbf{if } L{=}R \textbf{ then } ok \\
& \textbf{else } \ (\Sigma L'', R'', t''\cdot\ (L''{=}L) \times (t''{=}t) / 4 \times (L''{=}R'') \times (L'{=}L'') \times (R'{=}R'') \times (t'{=}t''{+}1)) \\
& \quad + (\Sigma L'', R'', t''\cdot\ (L''{=}L) \times (t''{=}t) / 4 \times (L''{\ne}R'') \times (L'{=}R') \times (t'{>}t''{+}1) \times 2^{t''+1-t'}) \\
& \quad + (\Sigma L'', R'', t''\cdot\ (R''{=}R) \times (t''{=}t) / 4 \times (L''{=}R'') \times (L'{=}L'') \times (R'{=}R'') \times (t'{=}t''{+}1)) \\
& \quad + (\Sigma L'', R'', t''\cdot\ (R''{=}R) \times (t''{=}t) / 4 \times (L''{\ne}R'') \times (L'{=}R') \times (t'{>}t''{+}1) \times 2^{t''+1-t'}) \textbf{ fi}
\end{aligned}
$$

All the L and R variables, with any number of primes, are two-valued.
In the **else** context, $L{\ne}R$. So in the second line, R'' can only be R .
Similarly, in the last line, L'' can only be L .
So we can one-point all double-primed variables.

$$
\begin{aligned}
=\;& \textbf{if } L{=}R \textbf{ then } ok \\
& \textbf{else } \ (L'{=}R'{=}L) \times (t'{=}t{+}1) / 4 \\
& \quad +\ (L'{=}R') \times (t'{>}t{+}1) \times 2^{t+1-t'} / 4 \\
& \quad +\ (L'{=}R'{=}R) \times (t'{=}t{+}1) / 4 \\
& \quad +\ (L'{=}R') \times (t'{>}t{+}1) \times 2^{t+1-t'} / 4 \textbf{ fi}
\end{aligned}
$$

In the **else** context, $L{\ne}R$, so exactly one of the first and third lines is nonzero.
The second and fourth lines are identical, and can be added.

$$
\begin{aligned}
=\;& \textbf{if } L{=}R \textbf{ then } ok \\
& \textbf{else } \ (L'{=}R') \times (t'{=}t{+}1) / 2 \\
& \quad +\ (L'{=}R') \times (t'{>}t{+}1) \times 2^{t+1-t'} / 2 \textbf{ fi}
\end{aligned}
$$

Where $t'{=}t{+}1$, dividing by 2 is the same as multiplying by $2^{t-t'}$.
And simplify the last line.

$$
\begin{aligned}
=\;& \textbf{if } L{=}R \textbf{ then } ok \\
& \textbf{else } \ (L'{=}R') \times (t'{=}t{+}1) \times 2^{t-t'} \\
& \quad +\ (L'{=}R') \times (t'{>}t{+}1) \times 2^{t-t'} \textbf{ fi}
\end{aligned}
$$

now combine the two terms

$$
\begin{aligned}
=\;& \textbf{if } L{=}R \textbf{ then } ok \textbf{ else } (L'{=}R') \times (t'{>}t) \times 2^{t-t'} \textbf{ fi} \\
=\;& loop
\end{aligned}
$$

Now we put the initialization together with the loop distribution to calculate the final state distribution.

$$
\begin{aligned}
& (t'{=}t)/4.\ \ loop \\
=\;& (L'{=}R') \times (t'{\ge}t) \times 2^{t-t'-1}
\end{aligned}
$$

omitting several steps

The average value of t' is

$$
\begin{aligned}
& (L'{=}R') \times (t'{\ge}t) \times 2^{t-t'-1}.\ \ t \\
=\;& t{+}1
\end{aligned}
$$

omitting several steps

On average, Mr.Bean draws the initial two socks plus one more sock from the drawer.