00314

The dynamic programming equations are

$$\begin{aligned}
W_i &= 0.95 \max\left\{2^{i-1}, \frac{9}{9+i}W_{i+1}\right\} + 0.05 \max\left\{2^{i-1}, \frac{6}{6+i}W_{i+1}\right\} \\
V_i &= 0.95 \max\left\{2^{i-1}, \frac{9}{9+i}V_{i+1}, \frac{10}{10+i}W_{i+1}\right\} \\
&\quad + 0.05 \max\left\{2^{i-1}, \frac{6}{6+i}V_{i+1}, \frac{7}{7+i}W_{i+1}\right\}
\end{aligned}$$

$i = 1, \ldots, 9$, where as boundary conditions we take $V_{10} = W_{10} = 2^9$.

From these we find $W_9 = 2^8$ and $V_9 = (19/20)(10/19)2^9 + (1/20)2^8 = (21/20)2^8$.

If Q_8 is easy the contestant can either retire (reward 2^7), attempt to answer (expected reward $(9/17)(21/20)2^8$), or phone a friend and then answer (expected reward $(10/18)2^8$). Now a short calculation verifies that $(9/17)(21/20) > (10/18)$, so the best option is to answer without phoning a friend.

If the number of potential question is to be unlimited this is a case of an positive programming over the infinite horizon (i.e., maximizing positive rewards). It is a theorem for positive programming that *if a policy has a value function that satisfies the dynamic programming equation, then that policy is optimal.*

So consider a policy in which the contestant retires whenever she has answered 9 or more questions. This policy has $V_i = W_i = 2^{i-1}$ for all $i > 9$. Easily observe that these values satisfy the dynamic programming equation for all $i > 9$. Therefore, by the theorem quoted above, it is optimal to retire once 9 questions have been correctly answered. So far as optimal play is concerned, the contestant will never wish to attempt $Q_{10}, Q_{11}, \ldots$.