The 'find the oldest' problem

A teacher knows that all n pupils in her class were born in 1980. She wishes to identify the oldest pupil, using a policy that will minimize the expected number of questions that she has to ask. She may only ask 'yes' and 'no' questions, but these may be of any form whatsoever, e.g.,

- Were you born before 4pm on February 14 or after 9am on June 10?
- Were you born in a month ending in 'y'?

The teacher wants to know

(a) Can she restrict her questions to ones of the form "Were you born before x?"

(b) As a function of n, what is the minimal expected number of questions required?

This is difficult problem, and an answer was not known until about five years ago.

There is a DP equation of the form

$$F(t,x) = \min_{Q_t}\{1 + EF(t+1, x + (Q_t; z_t))\}$$

where Q_t is the tth question, z_t is a vector of the t answers and $F(Q_1,\dots,Q_t; z_1,\dots,z_t,t) = 0$ if the answer to these t questions contains enough information to deduce the answer.

But this gives little clue. If fact the answer to (a) is yes. When $n = 2$ the optimal first question is, "Were you born before the midpoint of the year?". If just one pupil answer 'yes' we are done. Otherwise the problem starts over with both pupils in the same half year. Hence $F = 2$.

The answer to (b) is the the expected number of questions required is < 3 for all n. This seems surprising at first. But notice that if she asks "Were you born in the first $1/n$th of the year?" then the expected number of people answering 'yes' is one. If none answer 'yes' then the problem is as before. Otherwise it is likely that only 2 or 3 will answer 'yes' and then this small problem can be solved.