Let $q \in [0,1]^N$ be a fixed vector where $q_i = P(X_i = 1)$. In each time-step t upto a known end point T:

1. The learner chooses an $I_t \in \{0, \dots, N\}$ and $J_t \in \{0, 1\}$, setting $X_{I_t,t} = J_t$. Selecting $I_t = 0$ corresponds to taking the nothing action $do()$ and just observing.
2. For $i \neq I_t$, $X_{i,t} \sim Bernoulli(q_i)$
3. The learner observes $X_t = [X_{1,t} ... X_{N,t}]$
4. The learner receives reward $Y_t \sim \text{Bernoulli}(r(X_t))$, where $r : \{0,1\}^N \to [0,1]$ is unknown and arbitrary.

The causal structure gives us:

$$
\begin{aligned}
P(Y|do(X_i = j)) =& P(Y|X_i = j) \\
=& P(Y|do(X_a = 1), X_i = j)q_a + P(Y|do(X_a = 0), X_i = j)(1 - q_a)
\end{aligned}
$$

At each timestep, observing will reveal the reward for half the arms.