APPENDIX

PROOF OF LEMMA 1

Lemma. Assume the user is looking for a single entity of interest by exploring the ranked list of entities one-by-one from top to bottom, until this entity is found. Expected user effort then is minimized if the entities are ranked in decreasing order of their effort-adjusted probability.

Proof: Without loss of generality, assume the entities are ranked in order $y_1, y_2, \ldots, y_n$. If $y_k$ is the correct result, the user would have to go through entities $y_1$ to $y_{k-1}$, which are all ranked higher, until finding $y_k$. The corresponding total effort is $\sum_{j=1}^{k-1} \phi_j$. Hence, we obtain the expected user effort as

$$\sum_{i=1}^{n} \Pr(Y = y_i | A_1, \ldots, A_k, \bar{y}_1, \ldots, \bar{y}_i, D) \cdot \sum_{j=1}^{i} \phi_j.$$  (6)

Assume for contradiction that ranking $y_1, y_2, \ldots, y_n$ minimizes expected user effort, but entities are not ranked in decreasing order of their effort-adjusted probabilities. Then there exists a pair of adjacent (in the ranking) entities $(y_k, y_{k+1})$ that is not correctly ranked by effort-adjusted probability, i.e., $P_k/\phi_k < P_{k+1}/\phi_{k+1}$. (To avoid clutter, we use $P_r$ as a shorthand for $\Pr(Y = y_i | A_1, \ldots, A_k, \bar{y}_1, \ldots, \bar{y}_i, D)$.) We show that swapping their positions will decrease expected effort, contradicting the assumption of minimality.

It is easy to see that swapping $y_k$ and $y_{k+1}$ in Eq. 6 changes expected user effort by $P_k/\phi_k - P_{k+1}/\phi_{k+1}$. This difference is negative because of $P_k/\phi_k < P_{k+1}/\phi_{k+1}$. Hence the initial ranking $y_1, y_2, \ldots, y_n$ did not minimize expected user effort, completing the proof of Lemma 1 by contradiction.

PROOF OF LEMMA 2

Lemma. Let $p$, $r$, and $q$ be conditions for attribute $X$, and let the corresponding entity probabilities obtained based on these conditions be $P = \Pr(Y | A = p, A_1, \ldots, A_k, \bar{y}_1, \ldots, \bar{y}_i, D)$, $Q = \Pr(Y | A = q, A_1, \ldots, A_k, \bar{y}_1, \ldots, \bar{y}_i, D)$, and $R = \Pr(Y | A = r, A_1, \ldots, A_k, \bar{y}_1, \ldots, \bar{y}_i, D)$. If $q = p + \alpha \cdot (r - p)$, $0 < \alpha < 1$, then $Q = P + \alpha \cdot (R - P)$.

Proof: For simplicity and without loss of generality, collinearity will be shown for an attribute $X$ with three possible values $x_1$, $x_2$, and $x_3$. Hence a condition on $X$ is a probability distribution $p \in \{(p_1, p_2, p_3) | p_1, p_2, p_3 \geq 0, p_1 + p_2 + p_3 = 1\}$. Since the third probability is determined by the other two (all have to add up to 1), we only need to consider a probability vector $(p_1, p_2)$.

Consider three different conditions for attribute $X$, expressed as probability vectors $p = (p_1, p_2)$, $q = (q_1, q_2)$, and $r = (r_1, r_2)$. Let the three vectors satisfy $q = p + \alpha \cdot (r - p)$ for some $0 < \alpha < 1$, i.e., $p, q,$ and $r$ are collinear conditions for attribute $X$. (Intuitively, $q$ lies on the line connecting $p$ and $r$). We show that the corresponding entity probabilities

$$P = \Pr(Y | A = p, A_1, \ldots, A_k, \bar{y}_1, \ldots, \bar{y}_i, D)$$

$$Q = \Pr(Y | A = q, A_1, \ldots, A_k, \bar{y}_1, \ldots, \bar{y}_i, D)$$

$$R = \Pr(Y | A = r, A_1, \ldots, A_k, \bar{y}_1, \ldots, \bar{y}_i, D)$$

have to be collinear as well, i.e., satisfy $Q = P + \alpha \cdot (R - P)$.

To see this, recall that each condition $a_i \in A_i$ is a distribution over the values of the corresponding attribute $X_i$. Hence the above probabilities are actually expectations over these combinations of $X$-values. Formally, $P$ and similarly $Q$ and $R$ is defined as

$$E_{X, X_1, \ldots, X_k} \Pr(Y | X, X_1, \ldots, X_k, \bar{y}_1, \ldots, \bar{y}_i, D).$$

Since expectations can be decomposed, we can equivalently write

$$E_X [E_{X_1, \ldots, X_k} \Pr(Y | X, X_1, \ldots, X_k, \bar{y}_1, \ldots, \bar{y}_i, D)].$$

Based on the definition of the expectation, we then obtain

$$P = p_1 \cdot g(x_1) + p_2 \cdot g(x_2) + (1 - p_1 - p_2) \cdot g(x_3)$$

$$Q = q_1 \cdot g(x_1) + q_2 \cdot g(x_2) + (1 - q_1 - q_2) \cdot g(x_3)$$

$$R = r_1 \cdot g(x_1) + r_2 \cdot g(x_2) + (1 - r_1 - r_2) \cdot g(x_3)$$

Since $q = p + \alpha \cdot (r - p)$, we can derive from Equation 8

$$Q = (p_1 + \alpha (r_1 - p_1))g(x_1) + (p_2 + \alpha (r_2 - p_2))g(x_2) + (1 - (p_1 + \alpha (r_1 - p_1)) - (p_2 + \alpha (r_2 - p_2)))g(x_3)$$

$$= (p_1g(x_1) + p_2g(x_2)) + \alpha (r_1g(x_1) + r_2g(x_2) + (1 - r_1 - r_2)g(x_3)) - (p_1g(x_1) + p_2g(x_2) + (1 - p_1 - p_2)g(x_3)).$$

Together with Equations 7 and 9, we then obtain the desired result that $Q = P + \alpha \cdot (R - P)$.