RC25227 (W1110-114) October 21, 2011
Mathematics

IBM Research Report

"Optimal" Node Ordering in Bayesian Belief Networks

Witichai Sachchmarga*, Lea Deleris, Bonnie Ray
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

* The work of Witi Sachchamarga was begun during a summer internship in the Business Analytics and Math Sciences department at IBM's T.J. Watson Research Center. The work was completed as part of the course requirements for the Ph.D. in Management Science and Engineering, Stanford University, under the supervision of Professor Ronald Howard.

Abstract

Eliciting probability distributions from experts has always been one of the most difficult challenges in building Bayesian Belief Networks. Most of the existing literature has focused on different tools to use during the elicitation process, or various cognitive biases to avoid. However, there are very few guidelines on the order in which nodes that should be presented to the experts during the elicitation process, or how experts react to different orders of nodes. There is also a lack of research on the potential use of Internet webpage as a new means for the probability elicitation process.

In this tutorial paper, we developed some models for determining the order in which the nodes are to be elicited based on various assumptions. We conducted a probability elicitation experiment with some graduate students using an online elicitation tool that we have developed in order to explore the implications of our node-ordering models and the potential of eliciting probability through an online webpage.

The experimental results indicate that 41% of the students experienced fatigue even during a 10-minute online probability elicitation survey, and that the node ordering in which the number of conditioning parents is increasing is the user-friendliest model out of the three different node-ordering models that we have created. It may also be possible to mitigate the difficulty of probability elicitation that arises from the high number of conditioning parents by changing the node ordering. These experimental results show that the selection of node ordering for the probability elicitation process cannot be taken for granted given a large Bayesian Belief Network, and further research on this subject is worth exploring.

Contents

1 Introduction

Bayesian Belief Networks [Howard and Matheson, 1984] are graphical models that represent probabilistic relationships among a set of chance variables. Bayesian Belief Networks (BBNs) have become more popular in decision analysis during the past decade because of the robust inference algorithms for updating probability distribution in the context of observed evidence.

One of the emerging practices for constructing a BBN is shown in the figure below. The elicitation engineer first inherits a BBN structure constructed by a network builder, as shown in the figure below. The elicitation engineer then interviews the domain experts and uses their beliefs as the probability input for that BBN [Meyer and Booker, 2001].

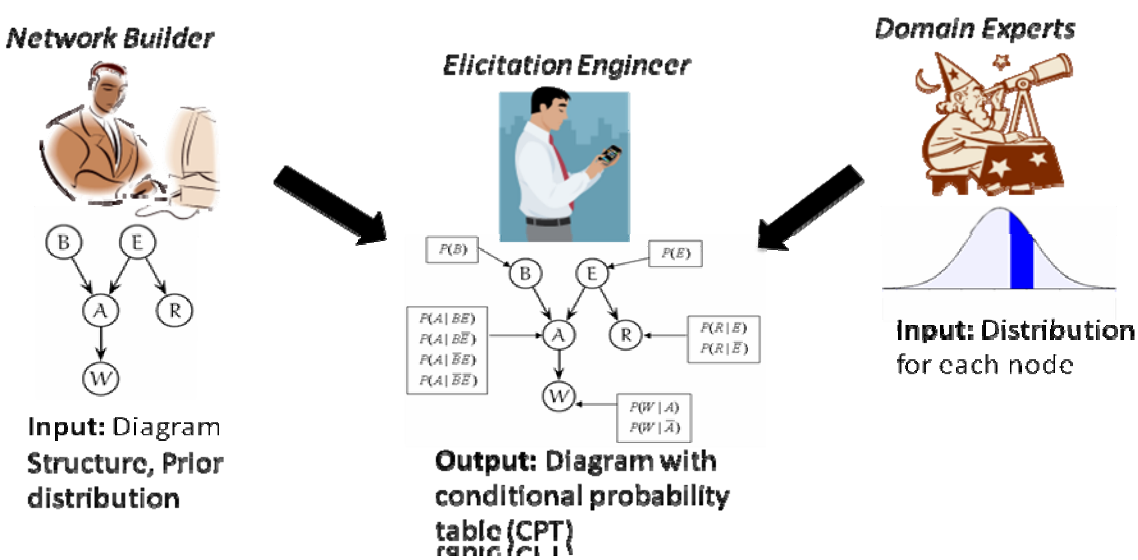


Figure 1: Process of building BBN involving Network Builder, Elicitation Engineer and Domain Experts

The workload is often divided up such that the network builder and elicitation engineer are not necessarily the same person, and the most difficult and time-consuming task is usually the elicitation of the probability distributions [Henrion, 1989; Druzdzel and van der Gaag, 1995]. Many scholars have recognized this problem, and a wide range of research has been conducted on how to improve the probability elicitation process with respect to using different tools [Wang and Druzdzel, 2000; Hope, Nicholson and Korb, 2002] or recognizing cognitive biases [Kahneman et al, 1982]. However, there are still many challenges for gathering probability distributions from experts in practice. For example, there are very few guidelines for the elicitation engineer regarding how she should arrange the order of the nodes to be elicited, given that each expert is responsible for multiple nodes in a given BBN [Meyer and Booker, 2001].

There has been some research on how to rearrange the node ordering for the second round of probability elicitation based on sensitivity analysis using the probability distribution elicited from the first round of elicitation [Castillo and Gutierrez 1997; van der Gaag, Renooji and Coupe, 2007]; however, none of it is focused on what the node ordering should be when the elicitation engineer meets the domain experts for the very first time. Similar research has been done in the context of decision diagrams [Lowell, 1994; Owen, 1978], but it also shares the same assumption, where certain fractiles (e.g., the 10^{th}, 50^{th} or 90^{th}) are already known from the first round of probability elicitation, and none of it has a focus on what the ordering should be for the very first round of the elicitation, where we do not have any fractiles from the experts to begin with. The current practice simply assumes that the top-down approach [Shachter]— which begins by eliciting the parent first and then works its way down to the children — is the best practice. While this particular ordering seems like a natural one, it still does not necessarily mean that it's the best ordering approach. We could not find any existing study that would explain why this would be the best node ordering, or what the potential disadvantages would be if the elicitation engineer were to deviate from this top-down approach. We're not aware of any other research on what the order of the nodes should be before any distribution has been elicited. We are also not aware of any existing research that describes how expert responses are affected by node ordering in terms of accuracy or friendliness. We tried to fill some of these research gaps in this tutorial paper. There are some tools such as knowledge maps [Howard, 1989] that can also be used to assist the node ordering; however, we have assumed in this paper that either the elicitation engineer does not have the liberty to modify the BBN that was given to her from the network builder or that the network builder has already used these tools when building the BBN.

Another research question that we explored in this tutorial paper is how experts would react to online probability elicitation tools. Experts are often too busy to meet for consecutive hours during the elicitation process, while using multiple experts around the world is often a logistical nightmare for the elicitation engineer. Building an online probability elicitation tool that allowed experts to log into the webpage on their own time to complete the elicitation process would mitigate some of these problems. Even with today's advances in technology, however, there are still few online elicitation tools available and even less research on how experts respond to them.

To summarize, we studied the following questions in this tutorial paper:

- For a particular BBN whose structure is known, in what order of nodes should the elicitation engineer try to elicit the distribution from the domain expert, based on the structure of the BBN?
- How would experts react to different orders of nodes?
- How would experts react to the online elicitation tool, and what are some possible improvements?

We feel that these are good research questions that have not been tackled before and that will provide for a good tutorial paper, as the problems could be easily expanded into a doctoral dissertation.

The remainder of the paper is organized as follows: in sections 2 and 3, we discuss two different models of node ordering based on various assumptions; in sections 4 and 5, we describe the design of our online elicitation tool and the elicitation experiment results using these tools with different node ordering; lastly, section 6 provides the conclusions of this paper and further research questions.

2 Node Ordering Model #1

To develop our first ordering model, we began by assuming that the expert would complete the entire elicitation process and that, as a result, all of the nodes would be elicited. We also assumed that the accuracy of the expert's answer for each node is approximately consistent. The "optimal" node ordering was then defined as the order that is the user-friendliest to the expert. The level of user-friendliness was measured during the experiment, based on feedback from experts after the elicitation process and elicitation rate.

In this section, we provide two heuristics for node ordering given the above assumptions. First, we propose that the elicitation engineer break the BBN down into separated layers, for three reasons: they make the ordering process much easier, as they break down the problem into smaller pieces; they can be used to clearly separate out the parents and children to guarantee that a child is never introduced before the parents; and they can provide nice stopping or breaking points for the elicitation process.

Rules for creating BBN layers

1. Parents must be at least one layer higher than their children.
2. Nodes with the same parents should be in the same layer, unless that violates rule #1
3. Nodes with the same children should be in the same layer, unless that violates rule #1 or #2

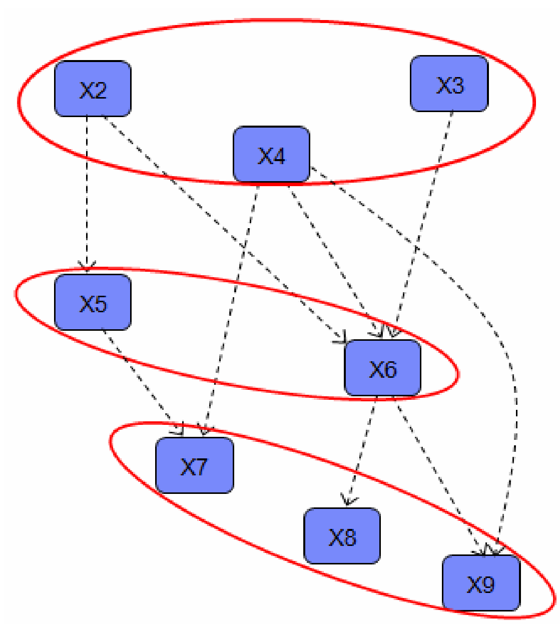


Figure 2: Creating BBN Layers

Algorithm for creating BBN layers

1. Create two sets: *Assigned* and *Unassigned*. Initial condition: ∀nodes are in the *Unassigned* set, *Assigned* set = empty
2. Find all nodes from the *Unassigned* set that do not have any parents, label them with layer number = 1 and move those nodes into the *Assigned* set
3. Identify all the nodes in the *Assigned* set that have the current highest layer number (let n = current highest layer) and find all of their children from the *Unassigned* set. Then do the following for those children nodes:
    i. Find the ones that do not have all of their parents already in the *Assigned* set and discard them
    ii. For the remaining undiscarded nodes, attach layer number = n+1 and move them into the *Assigned* set
4. Repeat step 3 until the *Unassigned* set is empty

2.1 Heuristic #1: Decreasing Parents

Heuristic #1 assumes that it would be easier for experts to tackle the most difficult questions first (the node that has many parents) because the experts would still be fresh at the start of the elicitation process.

Algorithm for Heuristic #1

1. Separate BBN into horizontal layers and assign each node to a layer
2. Starting from the top layer, order nodes in each layer using the ordering rule: {max children, max parents}

a. For each layer i, order each node from the highest number of children to the lowest
b. In event of a tie, order those tie nodes from highest number of parents to lowest
c. In the event of another tie, arrange those nodes into groups where all the nodes in the same group must share at least one parent
d. Randomly order the nodes within each group, starting from the group with the lowest number of members and moving to the highest

3. Repeat step 2 for the next layer down, until all layers are ordered

2.2 Heuristic #2: Increasing Parents

Heuristic #2 assumes that it would be easier for experts to tackle the simpler questions first (the node that does not have many parents), and then to slowly take on increasingly difficult ones later.

Algorithm for Heuristic #2

Use the same algorithm as Heuristic #1, with the following changes:

- In step 2, use the rule: {max children, min parents}
- In step 2d, randomly order the nodes within each group, starting from the group with the highest number of members and moving to the lowest

2.3 Comparison between two heuristics

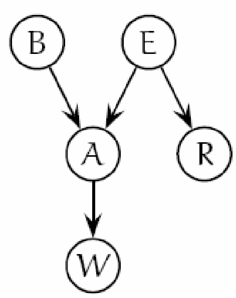
In this section, we provide the example of a different node ordering of three BBN structures using Heuristic #1 and Heuristic #2. These three BBN structures were selected from classic BBN literature; the summary of the ordering of each BBN using different heuristics is presented below. The objective of the analysis in this section was to determine whether or not the two heuristics would result in substantial differences in node ordering.

It can be seen in section 2.3.1 below that the two heuristics yield exactly the same ordering. However, this result is expected, as there are only five nodes in that BBN. Section 2.3.2 demonstrates that for a medium-sized BBN (eight nodes), the two heuristics will yield a moderate difference in node ordering. Section 2.3.3 shows that for a large BBN (38 nodes), the two heuristics will still yield the same ranking position for the majority of the nodes. However, the differences in position for some nodes are as large as five positions, which could lead to a completely different elicitation experience for experts.

Therefore, we conclude that for a small BBN (e.g., five nodes), the elicitation engineer should not have to worry about the different node ordering too much because the different approaches would probably still lead to the same node ordering. However, for a medium- to large-sized BBN, the question of node ordering is still worth exploring.

2.3.1 The Burglar-or-Earthquake Network

Mr. Holmes is working in his office when he receives a phone call from his neighbor Dr. Watson, who tells him that Holmes' burglar alarm has gone off. Convinced that a burglar has broken into his house, Holmes rushes to his car and heads for home. On his way, he listens to the radio, and in the news it is reported that there has been a small earthquake in the area. Knowing that earthquakes have a tendency to turn burglar alarms on, he returns to his work [Pearl, 2000].



- W: Phone call from Watson
- A: Burglary alarm
- B: Burglary
- R: Radio news
- E: Earthquake

Figure 3: BBN of the Burglar network

H1 Ordering

1. E
2. B
3. A
4. R
5. W

H2 Ordering

1. E
2. B
3. A
4. R
5. W

2.3.2 The Asia Network

A physician at a chest clinic wants to diagnose her patients with respect to three diseases based on observations of symptoms and possible causes of the diseases. The fictitious qualitative medical knowledge is the following.

The physician is trying to diagnose a patient who may be suffering from one or more of tuberculosis, lung cancer, or bronchitis. Shortness-of-breath (dyspnoea) may be due to tuberculosis, lung cancer, bronchitis, none of them, or more than one of them. A recent visit to Asia increases the chances of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis. The results of a single chest X-ray do not discriminate between lung cancer and tuberculosis, as neither does the presence or absence of dyspnoea. From the description of the situation it is clear that there are three possible diseases to consider (lung cancer, tuberculosis, and bronchitis). The three diseases produce three variables Tuberculosis (T), Cancer (L), and Bronchitis (B) with states no and yes. These variables are the targets of the reasoning and may, for this reason, be referred to as hypothesis variables. The diseases may be manifested in two symptoms (results of the X-ray and shortness-of-breath). The two symptoms produce two variables X ray (X), and Dyspnoea (D) with states no and yes. In addition, there are two causes or risk factors (smoking and a visit to Asia) to consider. The two risk factors produce variables Asia (A) and Smoker (S) with states no and yes [Lauritzen & Spiegelhalter, 1988].

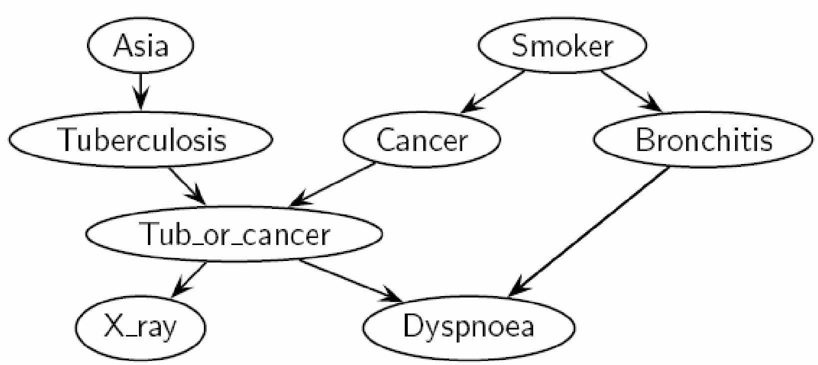


Figure 4: BBN of the Asian network

H1 Ordering

1. Smoker
2. Asia
3. Tuberculosis
4. Cancer
5. Bronchitis
6. Tub_or_cancer
7. Dyspnoea
8. X_ray

H2 Ordering

1. Smoker
2. Asia
3. Cancer
4. Bronchitis
5. Tuberculosis
6. Tub_or_cancer
7. X_ray
8. Dyspnoea

2.3.3 The Medical Network

A physician at a chest clinic wants to diagnose her patients with respect to three diseases based on observations of symptoms and possible causes of the diseases. The fictitious qualitative medical knowledge is the following [Veerle Coupe et al, 1999].

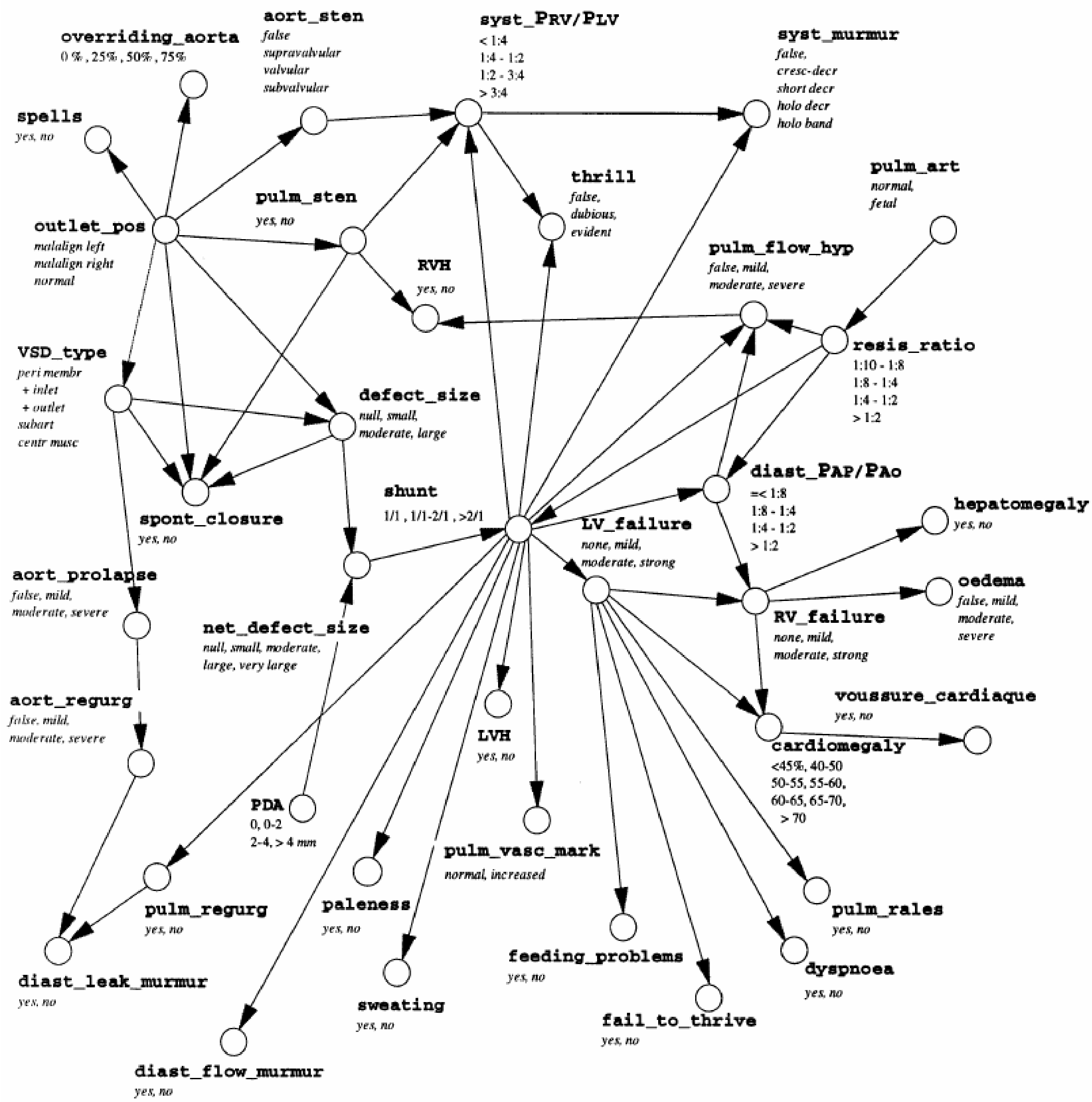


Figure 5: BBN of the Medical network

Table 1: H1 vs. H2 ranking for medical BBN

Node	H1	H2	ABS(H1-H2)
outlet_pos	1	1	0
pda	2	2	0
pulm_art	3	3	0
resis_ratio	4	6	2
pulm_sten	5	4	1
vsd_type	6	5	1
aort_sten	7	7	0
overriding_aorta	8	8	0
spells	9	9	0
defect_size	10	10	0
aort_prolapse	11	11	0
aort_regurg	12	13	1
net_defect_size	13	12	1
spont_closure	14	14	0
shunt	15	15	0
LV_failure	16	16	0
diast_pap_pao	17	18	1
syst_prv_plv	18	17	1
pulm_regurg	19	19	0
LVH	20	20	0
diast_flow_murmur	21	21	0
paleness	22	22	0
pulm_vasc_mark	23	23	0
sweating	24	24	0
Rv_failure	25	25	0
pulm_flow_hyp	26	26	0
dyspnoea	27	30	3
fail_to_thrive	28	31	3
feeding_problems	29	32	3
pulm_rates	30	33	3
diast_leak_murmur	31	29	2
syst_murmur	32	27	5
thrill	33	28	5
cardiomegaly	34	34	0
hepatomegaly	35	36	1
oedema	36	37	1
rhv	37	35	2
voussure_card	38	38	0
		Average = 0.9474	

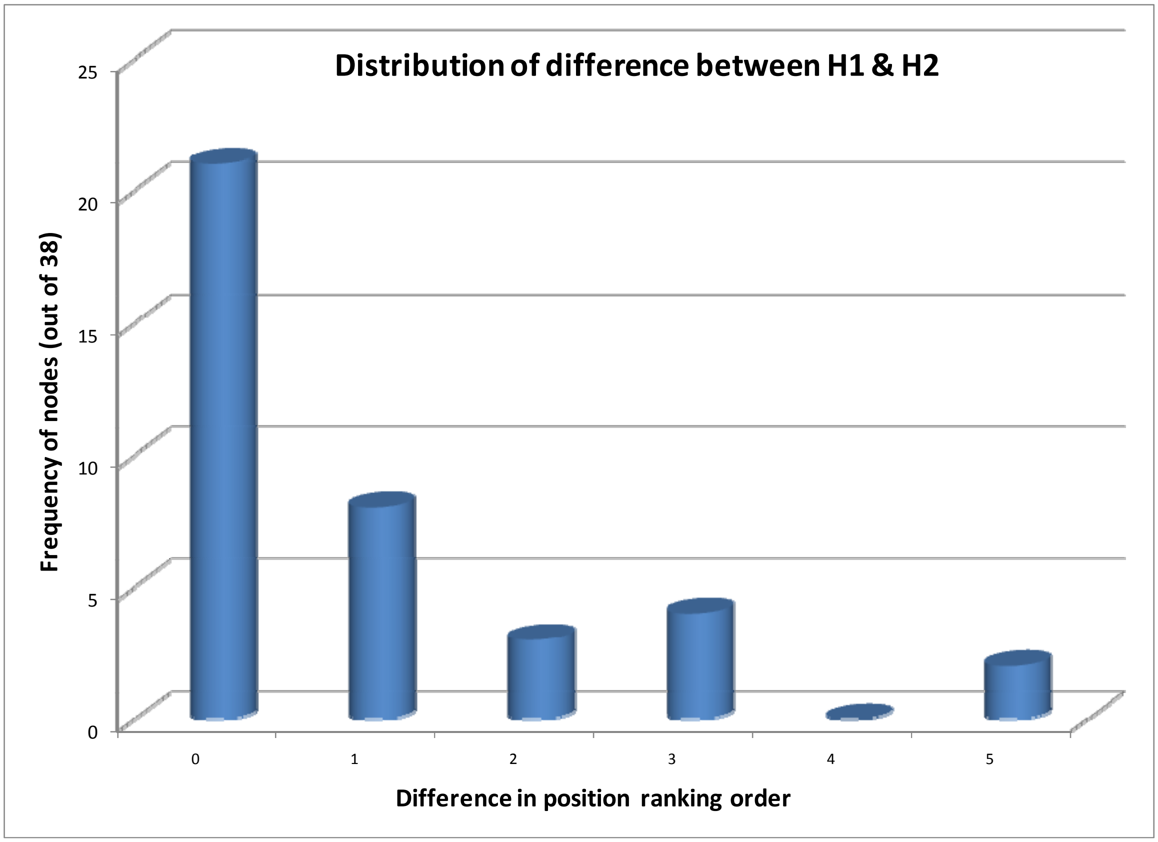


Figure 6: Frequency of differences between H1 & H2 for Table1

3 Node Ordering Model #2

To develop our second ordering model, we first assumed that the expert is an extremely busy individual and could quit the elicitation process at any time. We also assumed that experts are easily bored and tired, causing a significant decrease in the accuracy of their answers with each successive survey question. Therefore, the idea was to ignore the user-friendliness attribute in the node ordering approach outlined in section 2 and to instead order the elicitation node in terms of how important these nodes are to the probability distribution of the node in which we are truly interested (e.g., the revenue or profit node).

The dynamic "optimal" ordering would first identify which node should be elicited — assuming there is only enough time to elicit just one node — which would depend on the network structure and the prior probability provided by the network builder. The process would then be repeated under the same assumptions, but conditioned on the previous solution (e.g., given that node X will be elicited first, which node should be elicited next, given that there is only time to elicit one more node).

Ideally, one would derive a conditional variance for each node that has yet to be elicited and pick the next node to be elicited as the one that would yield the lowest conditional variance. However, for a BBN that is not very small, deriving the conditional variance is cumbersome. Therefore, a simulation analysis would be a better approach at this point. The simulation steps with reference to a BBN are shown in the figure below.

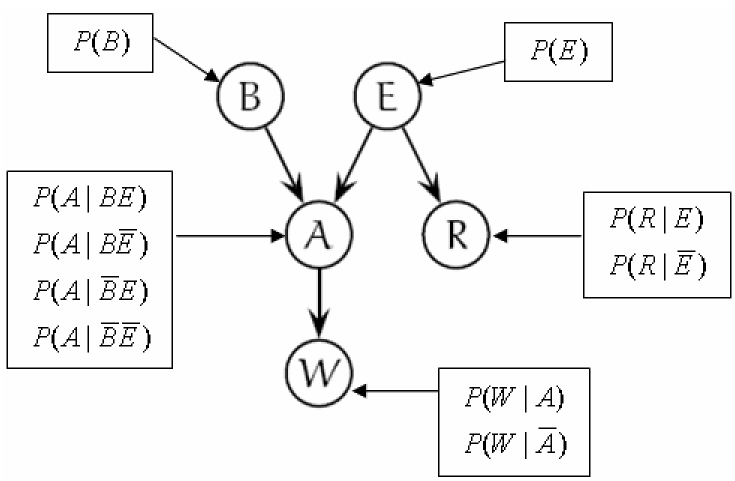


Figure 7: Example of BBN used in the outline of our simulation

Let W be the interested node (e.g., the profit node):

1. If a node has a parent, split up that node into 2^x sub-nodes (x is the number of parents); for example, node A in Figure 1 has two parents and therefore it is split into A_1, A_2, A_3 and A_4, as shown inside the box

2. Randomly generate 10 initial probability inputs from uni(0,1) for each question of the conditional probability table and call this the "initial probability"; these numbers will represent expert #1's true belief to be elicited inside the boxes in Figure 7.
3. Find the impact of question X_i by varying the input of all the other questions (except X_i) from 0 to 1 (or from 0.01-0.99) with specified step size (e.g., 0.1), and record the interval of P(W) as the measure for question X_i
4. Pick the question X_i with the smallest interval of P(W) to be asked first, then instantiate that question with the initial probability and repeat the above step again

After all X_i have been ranked, the set $\{X_i\}$ will give us the optimal order for expert #1. Generating 10 initial probability inputs again would yield expert #2's true belief. This process should be repeated. If n simulations were analyzed, this would be the equivalent of having sampled n random experts and developed the optimal question ranking for each expert.

3.1 Simulation Results

First, it turned out that varying the probability of other inputs in the conditional probability table (CPT) from 0 to 1 made it inconsequential which node we picked first. This surprising result can be explained due to the possibility that the CPT is filled with all zeros or ones. This is rather unreasonable, however, because when the probabilities are at zero or one, the nodes become deterministic and there is no reason to elicit their distribution any longer. Therefore, it is much more reasonable to focus on the simulation results that allow all other nodes that haven't been elicited yet to vary from [0.01, 0.99] rather than [0,1], as in step 3 of the previous section. We'll complete the explanation of our dynamic approach by applying our simulation two different BBNs, as shown below.

Three nodes in serial

The simulation was first done on three simple nodes in serial in order to study how the interval of the interested node changed. We defined the interval as follows: if, after eliciting node x, the probability of the interested node, p(C), could be in the interval [a, b], then we would define the length of the interval as (b-a). We chose to elicit the node that yielded the smallest p(C) interval.

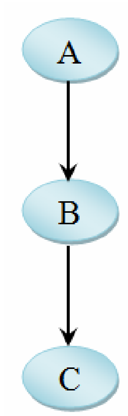


Figure 8: Three serial nodes, with node C assumed to be the interested node for elicitation

Table 2: 5 sample of simulation results for serial nodes

Trial #						Interval legth of p(C)				
1	C|B	C|~B	B|A	B|~A	A	0.97	0.09	0.09	0.00	0.00
2	C|~B	C|B	B|~A	A	B|A	0.97	0.29	0.28	0.07	0.00
3	C|B	C|~B	B|~A	A	B|A	0.97	0.64	0.64	0.13	0.00
4	C|B	C|~B	B|A	A	B|~A	0.97	0.14	0.14	0.03	0.00
5	C|B	C|~B	B|A	B|~A	A	0.97	0.30	0.30	0.03	0.00

The above table shows five samples of simulation results that actually give the order of the questions to be elicited for each expert. The table on the right shows the change in the interval p(C) after each question has been elicited. Initially, the interval length of p(C) is equal to one, as we have not elicited any part of the BBN and therefore only know that the p(C) is somewhere in the interval [0, 1]. The result of trial #1 (expert #1) tells us that for this particular expert, we should first ask for p(C|B) and that after eliciting that probability, the interval length of p(C) would reduce from 1 to 0.97. Then we should ask for p(C|~B), which would further reduce the interval of p(C) to 0.09. The last column should be all zeros because, at that point, all the probabilities would have been elicited and p(C) would simply be a single number with zero interval length. It should be noted that the interval length will always reduce to 0.97 after asking for the first conditional probability in this example, and the latter interval length will depend on the initial probability drawn for each expert. The simulation results clearly illustrates that the optimal ordering for this BBN structure would be to elicit node C first, then node B, and then finally node A.

Burglar Network

The simulation was also done on the burglar network shown in section 2.3.1. A sample of the simulation results is shown below.

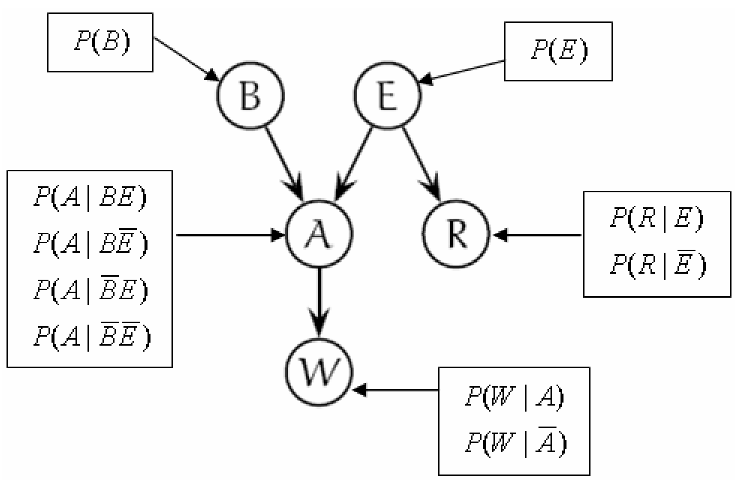


Figure 9: Burglar network used for simulation

Table 3: 5 sample of simulation results for Burglar network

Trial #										
1	W|A	W|~A	A|BE	A|B~E	B	E	A|~B~E	A|~BE	R|E	R|~E
2	W|A	W|~A	A|~BE	A|~B~E	E	A|B~E	A|BE	B	R|E	R|~E
3	W|A	W|~A	A|~BE	A|~B~E	E	A|B~E	A|BE	B	R|E	R|~E
4	W|A	W|~A	A|~BE	A|~B~E	B	A|B~E	E	A|BE	R|E	R|~E
5	W|A	W|~A	A|~BE	A|~B~E	B	A|BE	E	A|B~E	R|E	R|~E

Trial #	Interval Length of p(W)									
1	0.97	0.61	0.61	0.60	0.52	0.37	0.04	0.00	0.00	0.00
2	0.97	0.07	0.07	0.07	0.05	0.03	0.01	0.00	0.00	0.00
3	0.97	0.76	0.76	0.74	0.41	0.34	0.01	0.00	0.00	0.00
4	0.97	0.11	0.11	0.11	0.08	0.05	0.00	0.00	0.00	0.00
5	0.97	0.03	0.03	0.03	0.02	0.01	0.01	0.00	0.00	0.00

These simulation results actually give us a ranking for each column of the CPT that leads to the node ordering W, A, (B or E), R.

It should be noted that the R node is simply acting as a dummy node to check that our simulation is running correctly, since it should be obvious that the R node has no bearing on p(W).

4 Design of Online Elicitation Tool (*Risk Analytics Group, IBM Watson*)

With several node ordering models in hand, the next step was to put them to the test, which we did with a specially designed online elicitation tool.

The figures below show the step-by-step experience from the expert point of view when going through the online elicitation tool that we designed. Experts are first given a unique login and password so that each of their responses, along with time spent on each page, can be recorded easily. The elicitation tool should start with some simple elicitation exercises to familiarize the user with the tool and layout of the webpage, and to help the user overcome any other learning curve, before starting the actual elicitation process.

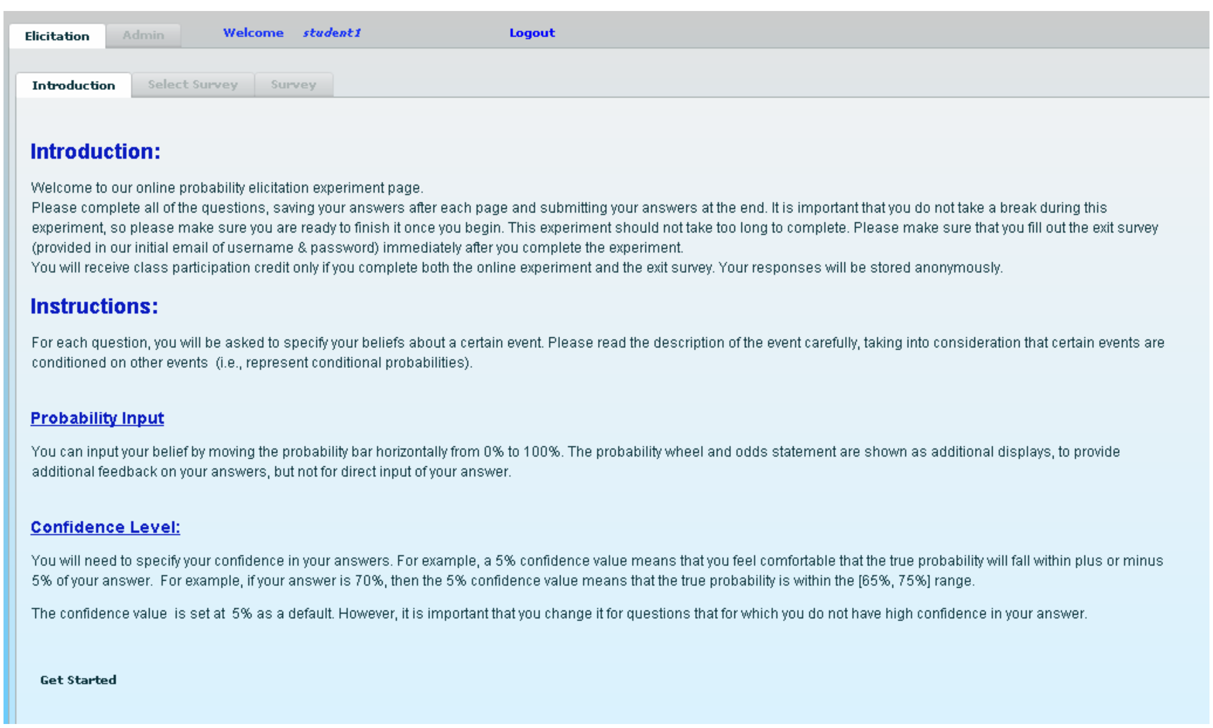


Figure 10: Introduction page explaining the general tool and outline of the elicitation page

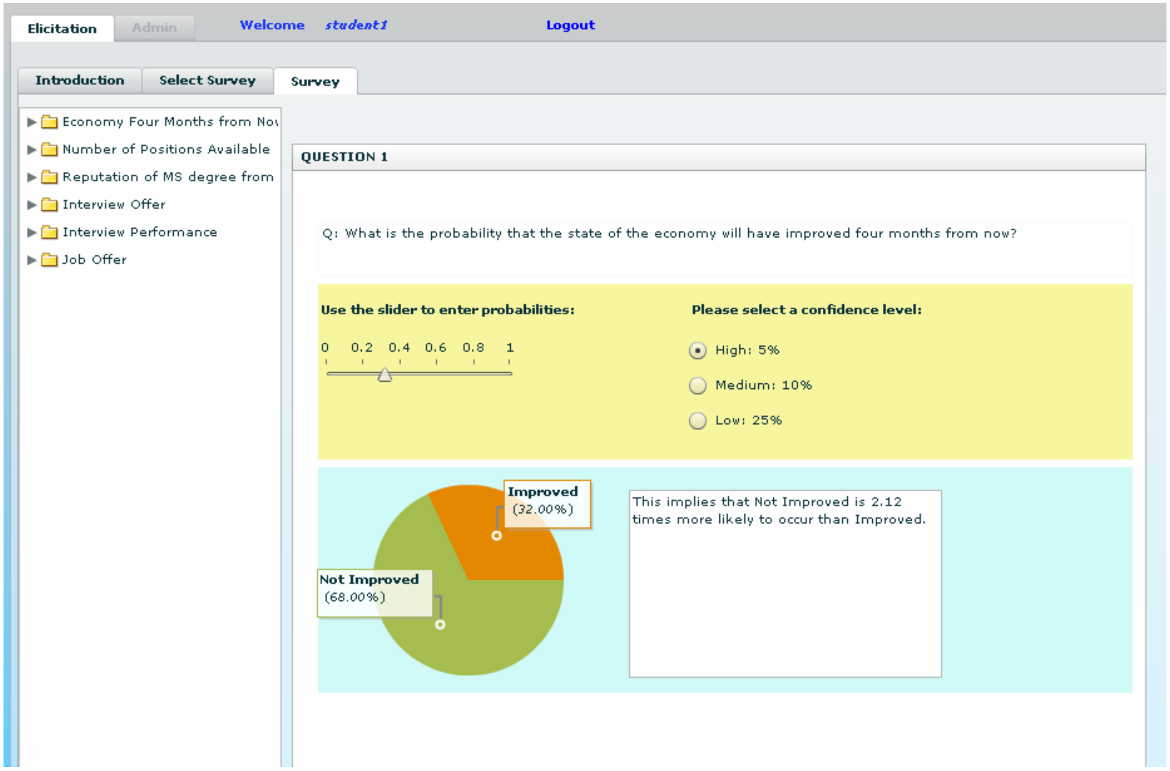


Figure 11: Sample page for eliciting a node with no parent

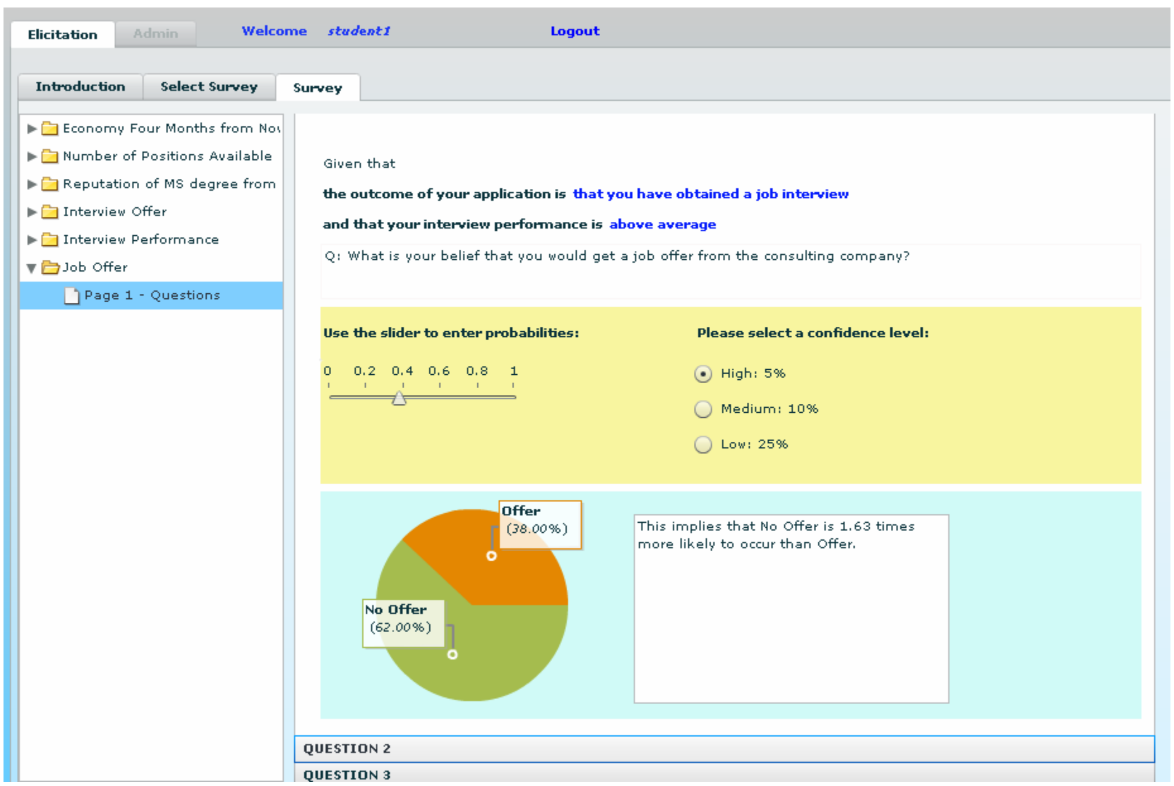


Figure 12: Sample page of eliciting a node with two parents

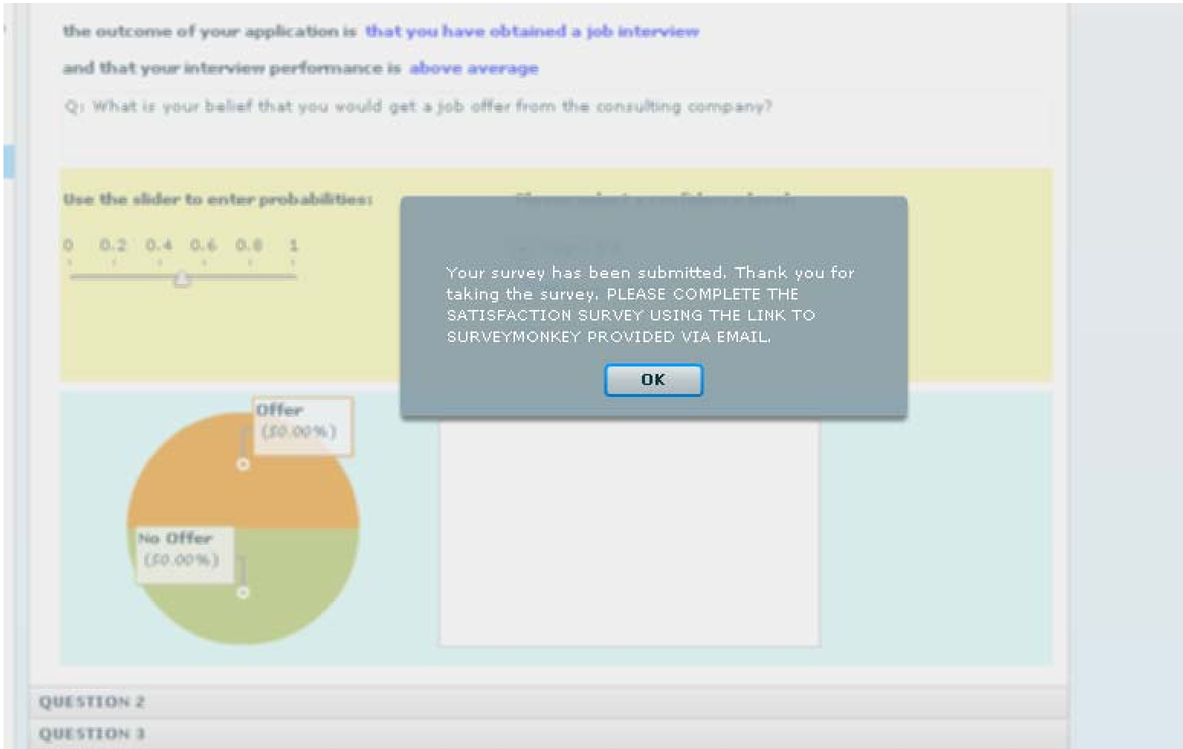


Figure 13: Sample page where the user reached the end of the elicitation process

5 Elicitation Experiment

We conducted an elicitation experiment to test our Heuristic #1, Heuristic #2, and dynamic ordering approaches. We asked students to log into a webpage elicitation tool similar to the one covered in section 4 and fill out a small BBN structure based on one of the randomly selected ordering approaches. We recorded the time for each elicitation, along with an exit survey to study the user-friendliness of the different approaches. We analyzed the data to answer the following questions:

- Which ordering approach is user-friendliest?
- Does the order of elicitation nodes really affect the overall experience of the experts?
- At what point do experts start to experience fatigue during the elicitation process and how does node ordering affect their elicitation rate or their confidence level in their answer?

We used the following BBN for this experiment:

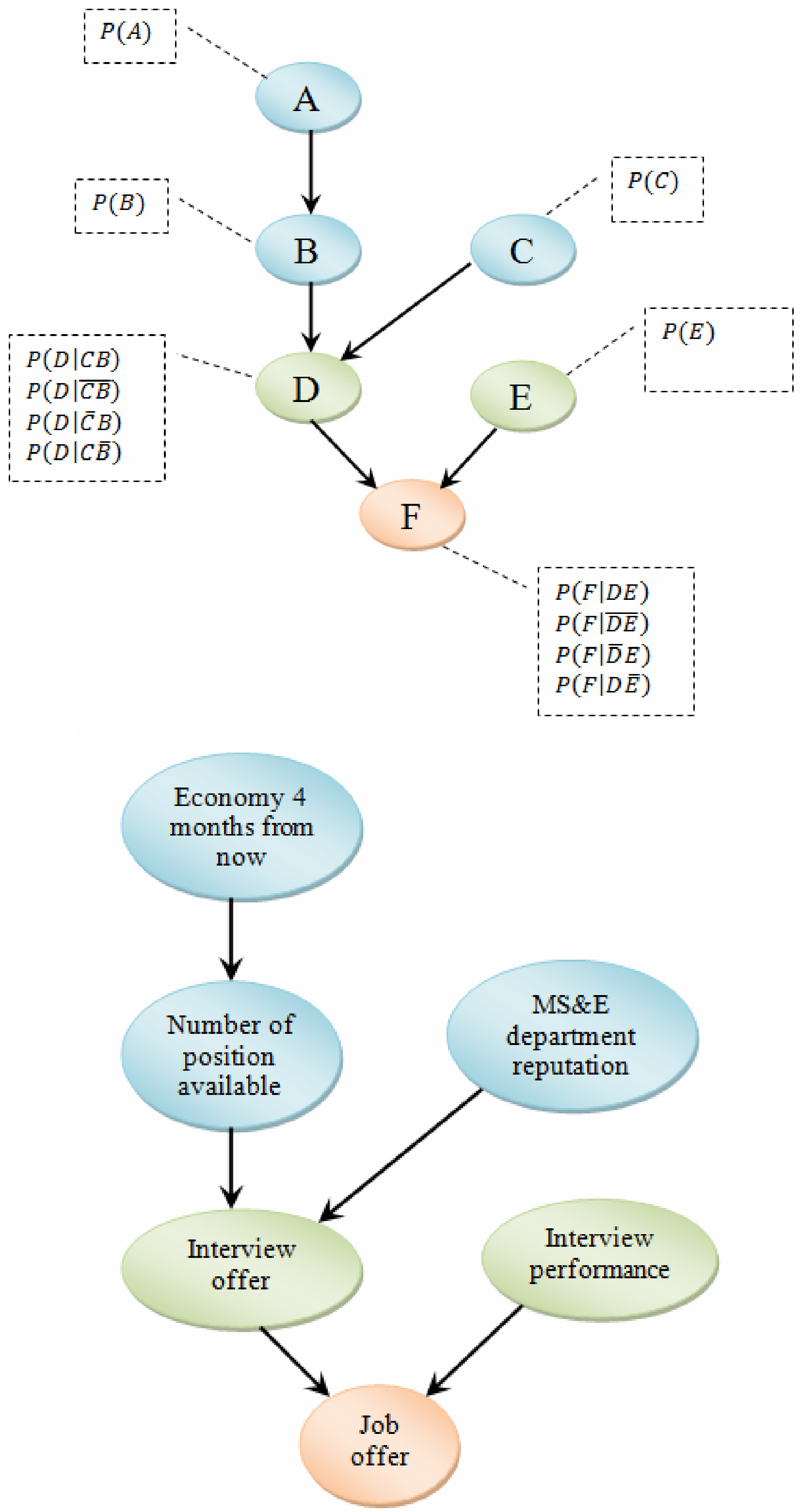


Figure 14: BBN used for the elicitation experiment with graduate students

We first asked students to imagine that they were about to graduate with a master's degree in management science and engineering at the end of the spring quarter of 2009 and that they would be actively seeking a position with a top consulting company within the next four months. We asked them to think about the ideal consulting company for which they would want to work. We assumed that the node of interest here was the last node ("Job offer from top consulting company"). We also assumed that the ideal consulting company would only hold one round of interviews in the form of a case study, and that the company would not inform applicants how many people they were looking to hire. The three different node orderings, as explained in section 2, are shown below.

Decreasing Parents

1. Economy 4 months from now
2. Number of position available
3. MS&E department reputation
4. Interview offer
5. Interview performance
6. Job offer

Increasing Parents

1. Economy 4 months from now
2. MS&E department reputation
3. Number of position available
4. Interview performance
5. Interview offer
6. Job offer

Dynamic

1. Job offer
2. Interview performance
3. Interview offer
4. MS&E department reputation
5. Number of position available
6. Economy 4 months from now

Node & states description

Economy 4 months from now: *Improve* vs*. Not Improve*

Improve = The economy 4 months from now will be better than what it is today

Not Improve = The economy 4 months from now will be the same or worst than what it is today

Number of position available: *Normal* vs*. Low*

Normal = The number of consultant that the company is hiring will be similar (or higher) to normal or regular year in the past.

Low = The number of consultant that the company is hiring will be lower than the normal or regular year in the past

MS&E department reputation: *Good* vs. *Bad*

Good = The consulting firm's opinion about a master in MS&E is generally good and competitive with an MBA or a master in Industrial and System Engineering.

Bad = The consulting firm's opinion about a master in MS&E is generally not as good and not as competitive as an MBA or a master in Industrial and System Engineering.

Interview offer: Interview vs. No Interview

Interview = You will receive an interview from the consulting company.

No Interview = You will not receive an interview from the consulting company.

Interview performance: *Great* or *Not Great*

Great = You will perform above average in your interview compare to the other candidates.

Not Great = You will perform the same or below average in your interview compare to the other candidates.

Job offer: *Offer* or *No Offer*

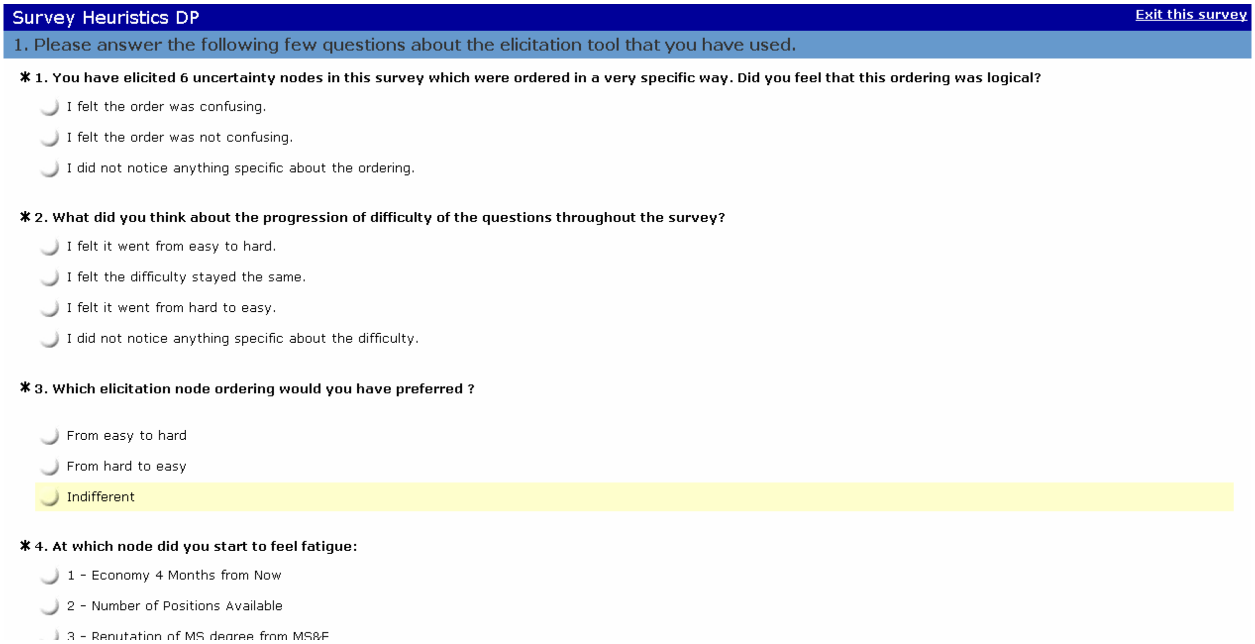
Offer= You will receive a job offer from the top consulting company.

No Offer= You will not receive a job offer from the top consulting company.

Exit Survey Page (*Risk Analytics Group, IBM Watson*)

Each of the three different node-ordering experiments led to an exit survey of its own. The students then completed the following 10 survey questions (see Appendix B for complete details):

1. How useful did you find the following probability elicitation display:
2. You have elicited 7 uncertainty nodes in this survey which were ordered in a very specific order. Do you think that the order of these nodes provided a good logical progression for the user?
3. What do you think about the progression of difficulty of the question throughout the survey?
4. Which elicitation node ordering would you have preferred between:
5. At which node did you start to feel fatigue:
6. Recall the node that you had the easiest time assessing. What was the main reason for this particular node to be the easiest for you?
7. Recall the node that you had the toughest time assessing. What was the main reason for this particular node to be the toughest for you?
8. How would you rate the online probability elicitation webpage that you've just completed in terms of?
9. What would you like to see made available that current isn't?
10. Any other comment?


Survey Heuristics DP
Exit this survey

1. Please answer the following few questions about the elicitation tool that you have used.

* 1. You have elicited 6 uncertainty nodes in this survey which were ordered in a very specific way. Did you feel that this ordering was logical?

- I felt the order was confusing.
- I felt the order was not confusing.
- I did not notice anything specific about the ordering.

* 2. What did you think about the progression of difficulty of the questions throughout the survey?

- I felt it went from easy to hard.
- I felt the difficulty stayed the same.
- I felt it went from hard to easy.
- I did not notice anything specific about the difficulty.

* 3. Which elicitation node ordering would you have preferred ?

- From easy to hard
- From hard to easy
- Indifferent

* 4. At which node did you start to feel fatigue:

- 1 - Economy 4 Months from Now
- 2 - Number of Positions Available
- 3 - Reputation of MS degree from MS&E

Figure 15: Example of the exit survey screen shot

5.1 Experimental Result Summary

The summary of the experiment, which involved 73 student participants, is as follows.

Heuristic #2 (increasing parents) had the lowest mean of elicitation time, followed by dynamic ordering, and lastly, Heuristic #1 (decreasing parents). The dynamic ordering has the lowest median of elicitation time. This is rather surprising, as we expected the dynamic ordering to be the most confusing, since it is purely based on the importance of each node and not on user-friendliness, as is the case with the two heuristics.

Table 4: Total elicitation time (seconds) for each ordering method

[redacted]

According to the exit survey, the number of students who found the node ordering to be confusing was 11, 2 and 0 for dynamic ordering, Heuristic #1 (decreasing parents) and Heuristic

#2 (increasing parents) respectively. This is not surprising, given that that the dynamic ordering initially elicits the node at the bottom and then sweeps upward.

This could also explain why dynamic ordering has the lowest median elicitation time. The experimental data show that there are four students who spent less than five seconds in eliciting their probability on a node from the dynamic ordering group, compared to none in the other two ordering groups. This statistic may suggest that dynamic ordering was too confusing for some students, which caused them to become careless in completing the elicitation exercise, leading to the lower median elicitation time.

Even though dynamic ordering has the lowest mean elicitation time, it also has the biggest standard deviation, along with the highest number of students who found it to be confusing from the exit survey. Therefore, we cannot suggest that it is the most user-friendly ordering. In fact, we would argue that the results of the experiment suggest that Heuristic #2 (increasing parents) is the most user-friendly ordering compared to the other two, since it has the lowest mean and standard deviation of the elicitation time. Heuristic #2 also has the best exit survey, since no student found the order to be confusing.

It can be seen in the table below that *high number of conditioning parents* and *less familiarity with subject matter* are two main reasons why students find eliciting probability in BBNs to be difficult. The results below also suggest that Heuristic #1 (decreasing parents) could mitigate the difficulties that arise from having a high number of parents compared to the other two orderings, since only 18% of students stated that *high number of conditioning parents* is the main reason for the difficulty in probability elicitation, compare to 68% and 46% in the other two orderings.

Table 5: Main reason why eliciting probability is difficult

[redacted]

About 41% of students did experience some fatigue during the elicitation process. This suggest that the elicitation engineer should consider fatigue as one of the key factor when designing her probability elicitation survey, even for the elicitation survey that will last for only 10 minutes.

Table 6: Number of students experiencing some fatigue during the elicitation process

	Decreasing Parents	Increasing Parents	Dynamics	**Total**
Fatigue	8	11	9	**28**
No Fatigue	14	11	15	**40**
Percent of Fatigue	36%	50%	38%	**41%**

We analyze the effect of different node orderings on the confidence level of probability elicited by applying a linear scoring rule. Confidence levels that are elicited as *high, medium* and *low* are scored using coefficient 2, 1 and 0 respectively. The confidence level for dynamic ordering was noticeably lower than the two heuristics after applying the linear scoring rule.

Table 7: Average score for each node ordering using linear scoring rule (higher score = higher confidence)

	Decreasing Parents	Increasing Parents	Dynamics
Mean	15.21	15.52	14.04

The plot below shows that the confidence level does not strongly increase as a function of time for any of the three node orderings.

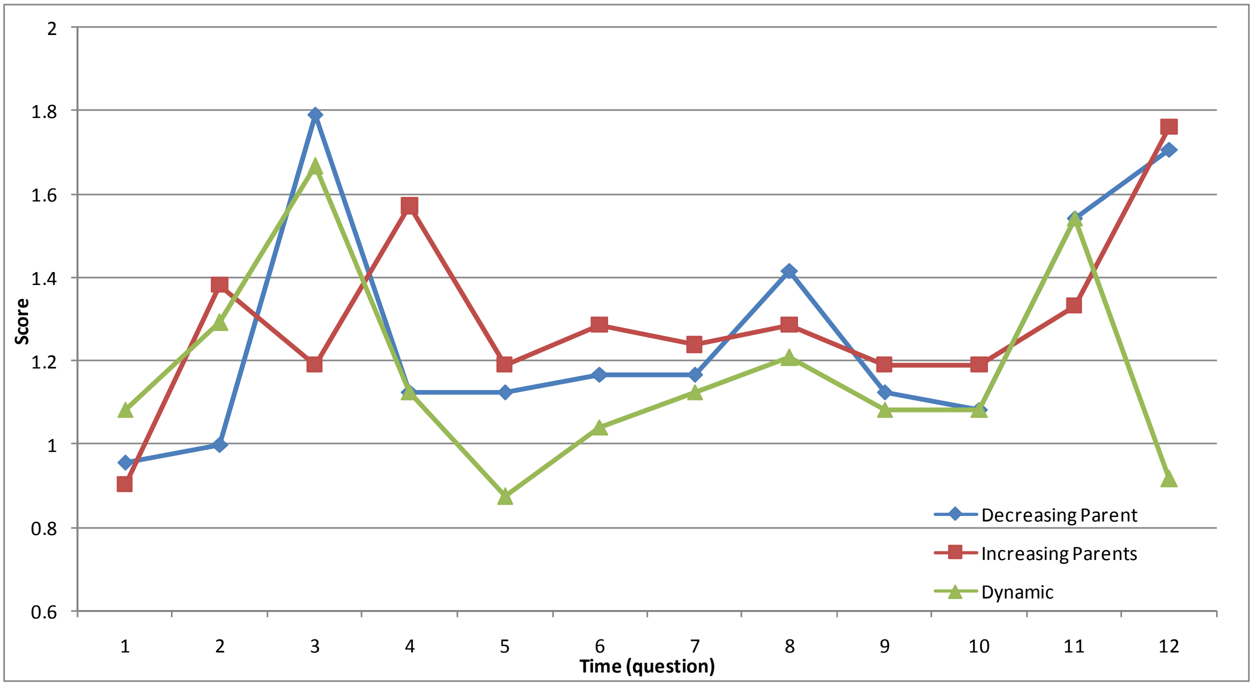


Figure 16: Change in confidence level as a function of time

However, the effect of different node orderings on the confidence level is more noticeable when we plot the confidence level as a function of a particular question (e.g., question #1 is *Economy 4 months from now*, question #12 is *Job offer*) as shown below. The plot shows that dynamic ordering does not perform very well compared to the other two orderings for each specific question.

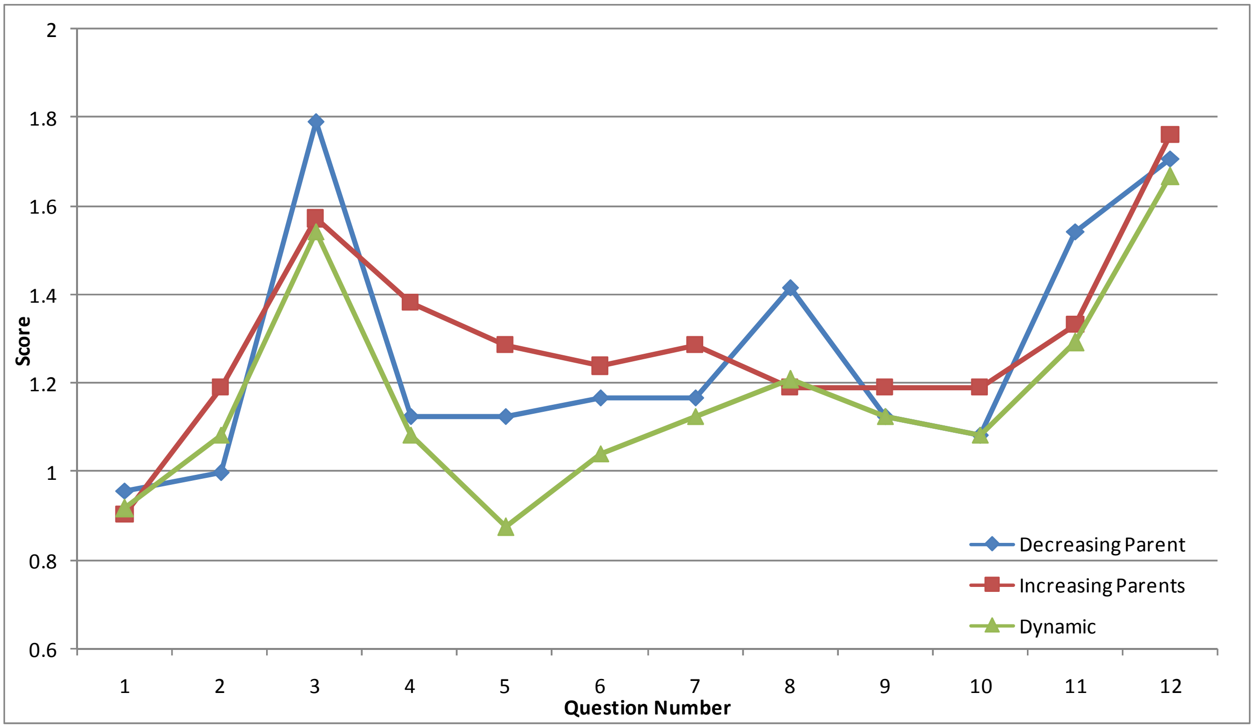


Figure 17: Change in confidence level as a function of subject matter

The table below shows the average effect of the number of conditional parents on the elicitation time and the confidence score of a node. The surprising results here indicate that the change in confidence level of a node is not really a function of the number of parents that it has.

Table 8: The effect of number of conditional parents to elicitation time and confidence score

Average	0 Parents	1 Parents	2 Parents
Elicitation time	45.19	58.30	125.70
Confidence score	1.09	1.09	1.18

In the end, most students were comfortable with the online elicitation tool and said that they wouldn't mind using it again. This suggests that there is a real potential to use Internet webpage as a new means for the probability elicitation process.

6 Conclusion

We have investigated some different node-ordering approaches for the probability elicitation process of Bayesian Belief Networks. We proposed three different ways of ordering the nodes using different assumptions.

The first two approaches to node ordering (heuristics #1 and #2) assume that experts will complete the entire elicitation process without any fatigue. Therefore, these two approaches are designed to provide domain experts with the most user-friendly order possible. The last ordering approach (dynamic) assumed that experts might quit at any time during the elicitation process and that the accuracy of their answers would decrease over time. Therefore, this approach simply ignored the user-friendliness aspect and ordered the nodes by their impact on the marginal distribution of a node of interest (e.g., the profit node).

We designed an online probability elicitation tool and conducted an experiment to see the impact of these node orderings based on the experience of 73 students. The general responses from the students were positive regarding the online tools, which suggests that there is real potential to use Internet webpage as a new means for the probability elicitation process. But this result might be biased, as most of the students have been exposed to the concept of probability elicitation in their decision analysis classes before.

The early experimental results also showed that Heuristic #2 (increasing parents) outperformed the other two ordering approaches in terms of user-friendliness, as Heuristic #2 has the best elicitation time, confidence level and exit survey responses. The results also indicate that using Heuristic #1 could mitigate the difficulties that arise from having a high number of parents, compared to the other two ordering approaches. The experiment shows that the elicitation engineer needs to worry about the fatigue factor, because 41% of the students did experience fatigue during the elicitation process even though the experiment was only designed for 10 minutes. In the end, the experimental results indicated that using different node orderings during a probability elicitation process could significantly impact the users' elicitation experience in terms of speed, friendliness, and confidence in answering.

6.1 Research Extension

The possible research extensions of this tutorial are as follows:

- The second approach of node ordering could be improved by modeling a distribution of the time at which experts quit the elicitation process. We could also create a model that would capture the decrease in accuracy of answers as a function of time.
- A hybrid of the two approaches could be created so that the elicitation engineer would have an ordering that is somewhat user-friendly to the expert but at the same time took into account an order that would have a large impact on a particular node.
- We could also investigate the ordering for nodes that are distributed with distributions other than Bernoulli.
- We could also use various distributions other than the uniform distribution to represent our prior beliefs for each node.
- Lastly, we could also investigate a scenario in which each expert has his/her own optimal node ordering, and the possibility of combining these individual orderings into a single global ordering.

References

Castillo, E., and Gutierrez, J., 1997, Sensitivity analysis in discrete Bayesian networks, *IEEE Transactions on System, Man and Cybernetic*

Chan, H., and Darwiche, A., 2004, Sensitivity analysis in Bayesian networks: From single to multiple parameters, In *Proceedings of the Seventeenth National Conference on Artificial Intelligence*

Coupe, V., Peek, N., Ottenkamp, J., Habbema, J., 1999, Using sensitivity analysis for efficient quantification of belief network, *Aritificial Intelligence in Medicine*

Druzdzel, M., and van der Gaag, L., 1995, Elicitation of probabilities for belief networks: Combining qualitative and quantitative information, *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*

Druzdzel, M., and van der Gaag, L., 2000, Building probabilistic networks: Where do the numbers come from?, Guest editors' introduction , *IEEE Transactions on Knowledge and Data Engineering*

Hesper, E., van der Gaag, L., Feelders, A., Loeffen, W., Geenen, P., Elbers, A., 2005, Bringing order into Bayesian network construction, *Proceedings of the 3rd international conference on Knowledge capture*

Hope, L., Nicholson, A., and Korb, K., 2002, Knowledge Engineering Tools for Probability Elicitation, *Technical Report, School of Computer Science and Software Engineering, Monash University.*

Howard, R., 1984, Influence diagram, *The Principles and Applications of Decision Analysis Vol. II*, Strategic Decision Group

Howard, R., 1989, Knowledge maps, *Management Science*

Kahneman, D., Slovic, P., and Tversky, A., 1982, Judgement under Uncertainty: Heuristics and Biases, *Cambridge University Press*

Koller, D., Friedman, N., 2002, *Manuscript in progress*

Lauritzen, S., Spiegelhalter, D., 1988, Local computations with probabilities on graphical structures and their application to expert systems, *Journal of Royal Statistical Society*

Lowell, D., 1994, Sensitivity to relevance in decision analysis, *Doctoral* d*issertation, Stanford University*

Meyer, M., Booker, J., 2001, Eliciting and analyzing expert judgment, *ASA-SIAM Series on Statistics and Applied Probability*

Owen, D., 1978, The concept of influence and its use in structuring complex decision problems, *Doctoral* d*issertation, Stanford University*

Pearl, J., 2000, Causality Models: Reasoning and Influence, *Cambridge Univ. Press*.

Schachter, R., Making Decisions in Intelligent Systems: Representing Uncertainty with Belief Networks and Influence Diagrams, *Manuscript in progress.*

van der Gaag, L., Renooij, S., Witteman, C., Aleman, B, and Taal, B., How to elicit many probabilities, 1999, *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*

van der Gaag, L., Renooji, S., and Coupe, V., 2007, Sensitivity analysis of probabilistic networks, *Studies in Fuzziness and Soft Computing*

Wang, H., and Druzdzel, M., 2000, User interface tools for navigation in conditional probability tables and elicitation of probabilities in Bayesian networks, *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*

Appendix A: Matlab Code

Available upon request.

Appendix B: Exit Survey Questions

1. How useful did you find the following probability elicitation display:
    a. Probability wheel (picture)
    b. Slider bar (picture)
    c. Odds statement (picture)

Answer Choices: {Very useful, Useful, Not Useful at All, Don't Use}

2. You have elicited 7 uncertainty nodes in this survey which were ordered in a very specific order. Do you think that the order of these nodes provided a good logical progression for the user?

Answer Choices: {I felt the order was noticeable in term of good logical progression, I felt the order was confusing, I did not notice anything specific about the order}

3. What do you think about the progression of difficulty of the question throughout the survey?

Answer Choices: {Easier to harder, Difficulty stayed the same, Harder to easier, I did not notice anything specific about the difficulty}

4. Which elicitation node ordering would you have preferred between:
    a. Starting with the easiest question, then work your way up to the hardest question
    b. Start with the hardest question, then work your way down to the easiest question

5. At which node did you start to feel fatigue:

Answer Choices: {name of node #1,2,3,4,5,6,7, No Fatigue }

6. Recall the node that you had the easiest time assessing. What was the main reason for this particular node to be the easiest for you?

Answer Choices: {Low number of conditioning parents, More familiarity with the subject matter, Low level of fatigue, Other (specify)}

7. Recall the node that you had the toughest time assessing. What was the main reason for this particular node to be the toughest for you?

Answer Choices: {High number of conditioning parents, Less familiarity with the subject matter, High level of fatigue, Other (specify)}

8. How would you rate the online probability elicitation webpage that you've just completed in terms of?
    a. Usefulness (e.g. ability to elicit probability from multiple experts in real business)
    b. Friendliness (e.g. easy for experts to use, easy to understand)

Answer Choices: {Excellent, Good, Average, Poor, Very Poor, Undecided}

9. What would you like to see made available that current isn't?
10. Any other comment?