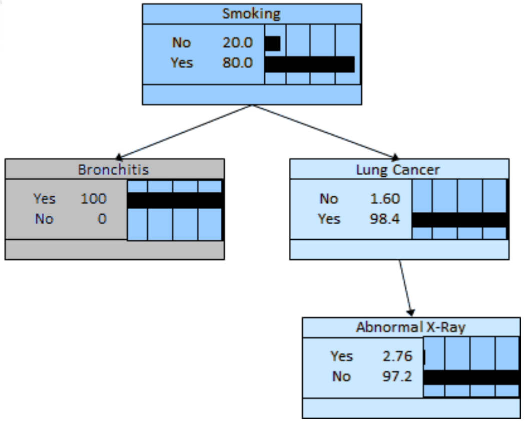
Bayes Nets (Bayesian networks)

Unleashing the power of an entirely new way to understand your data

Dr. Steven M. Struhl
smstruhl@gmail.com



What can Bayes Nets (Bayesian Networks) do?

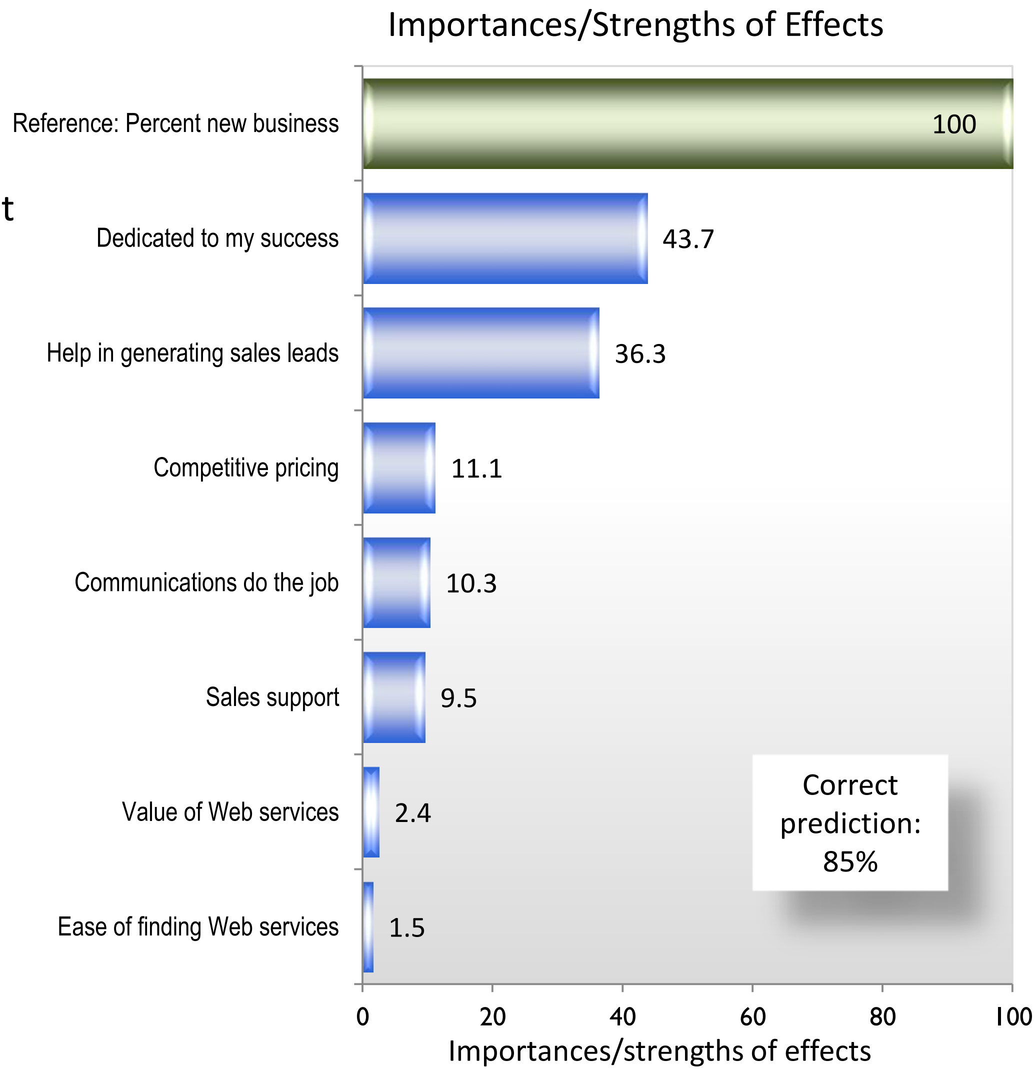
- They provide a whole new and highly powerful way to get inside data
- Practical applications are many:
  - **For top management, the best fix on what's important**
    - Finally seeing how questionnaire, database and Web variables **drive market share** and **revenues**
  - **For researchers, the story behind the data**
    - Determining how variables link together
    - Uncovering natural groupings in the data
    - Understanding how each variable affects both the target variable and all other independent variables
  - And many other applications, as we will see.
- Most importantly, **they work better** than what we have been using
  - Let's look at how, starting with the top management overview, then getting to the story behind the data—and then some explanation of their underpinnings.

The power of Bayes Nets: Questions vs. market share

- *Bayes Nets can tackle very large groups of variables, but let's start with a smaller set.*
  - **The situation:** Management has asked whether questionnaire questions in fact "drive" share
    - We have—in addition to share data (percent of new business awarded to the client)—a number of ratings from a satisfaction survey
    - We have already isolated these as most important:
      - Dedicated to my success
      - Helps in generating sales leads
      - Competitive pricing
      - Communications do the job
      - Sales support
      - Value of Web services
      - Ease of finding Web services
  - **The key questions:**
    - Exactly how important is each of these?
    - How much do they drive share?
    - What happens to share if we change each one?

For management: Bayes Nets reveal effects on share clearly

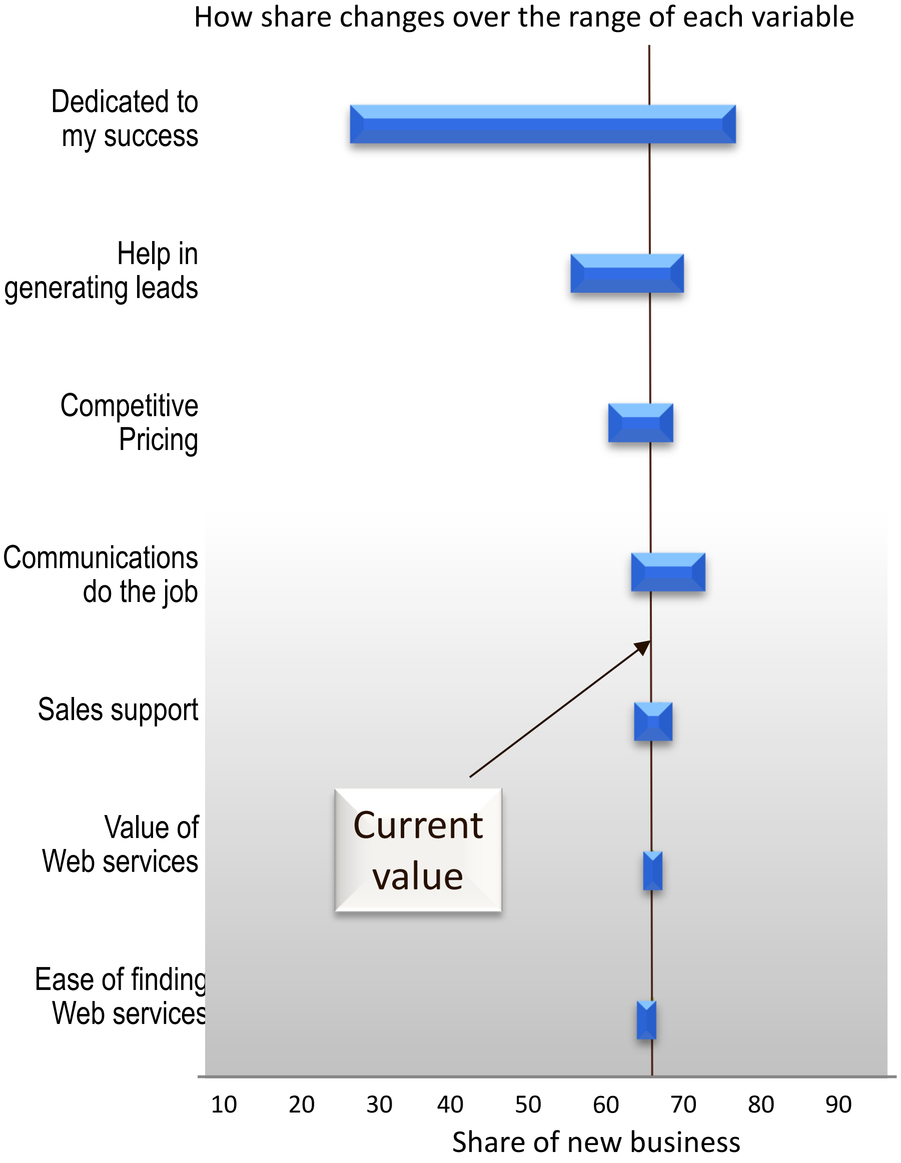
- We have confidence the model is accurate as it predicts share **85% correctly**
- With Bayes Nets correct prediction of share usually runs over 50% correct
  - This is very strong performance!
  - The best comparable regression-based model[1] predicted only 11% correctly—not a reliable indicator of what drives share
    - This low level for regression is typical when trying to forecast shares from questionnaire responses
- This is a familiar looking chart, but with a difference:
  - Effects are compared to what we could get if we could *reach in and directly move share*
- For instance, changing "dedicated to my success" has 43.7% of the effect of directly moving share
  - This is a powerful way to convey importances.



[1] *The best regression-based model was a partial least squares (PLS) path model*

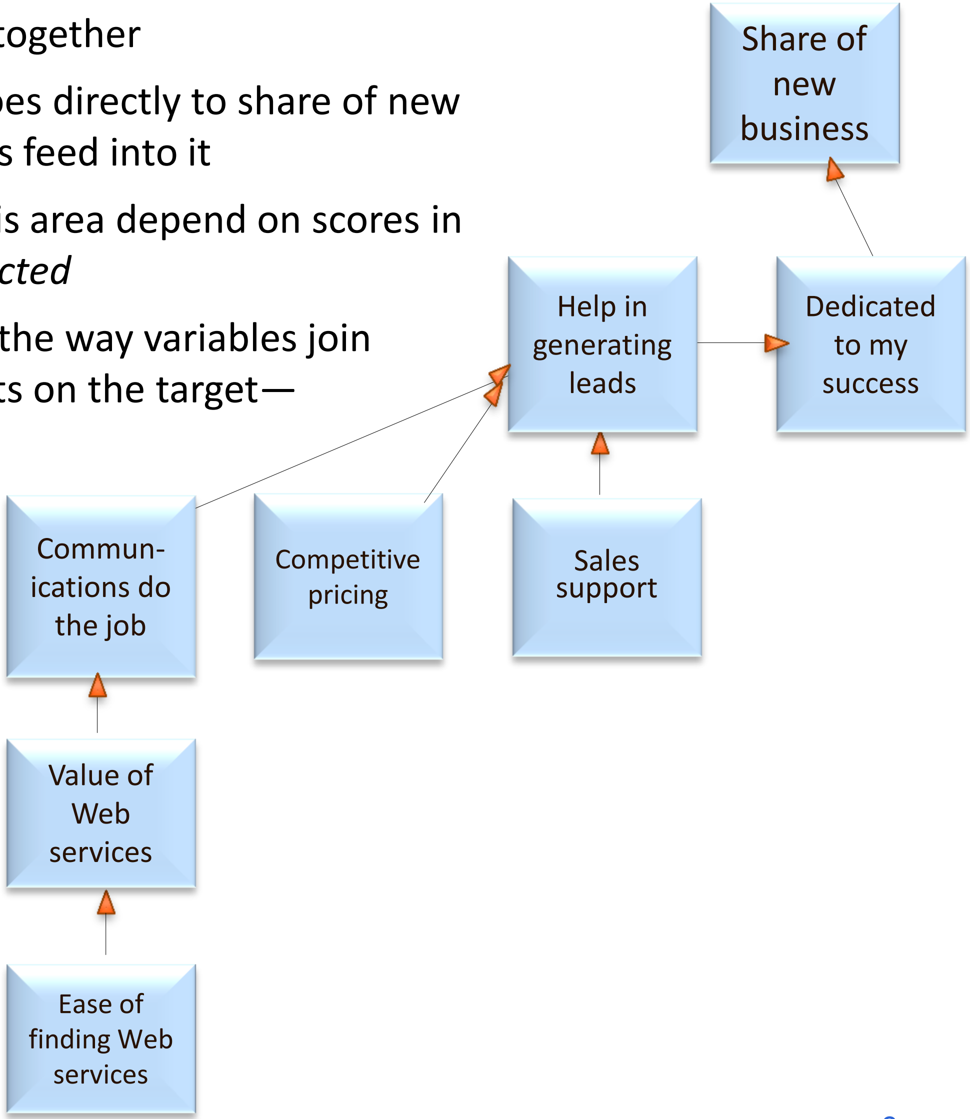
For management: Seeing exactly how each variable changes the target variable

- Here, results show precisely what happens as we change each variable over its range of ratings
  - There is real downside risk in the first two areas—when each rating is low, share of new business also is quite low
  - Poor scores in “dedicated to my success” in particular go with very low shares of new business
- Effects differ strongly by variable
  - e.g., with “Communications do the job,” there is more room to grow than to fall. Taking a real chance for gain here is less risky than in other areas
- This chart is so powerful it needs to be used with caution!
  - This does not mean that, i.e., just boosting “dedicated to my success” can drive share all the way up to 90%
  - It does show what would happen if we could do all that is needed to make these scores rise
  - The full Bayes Net diagram (to follow) will show that other scores are all connected and need to move so this score does.



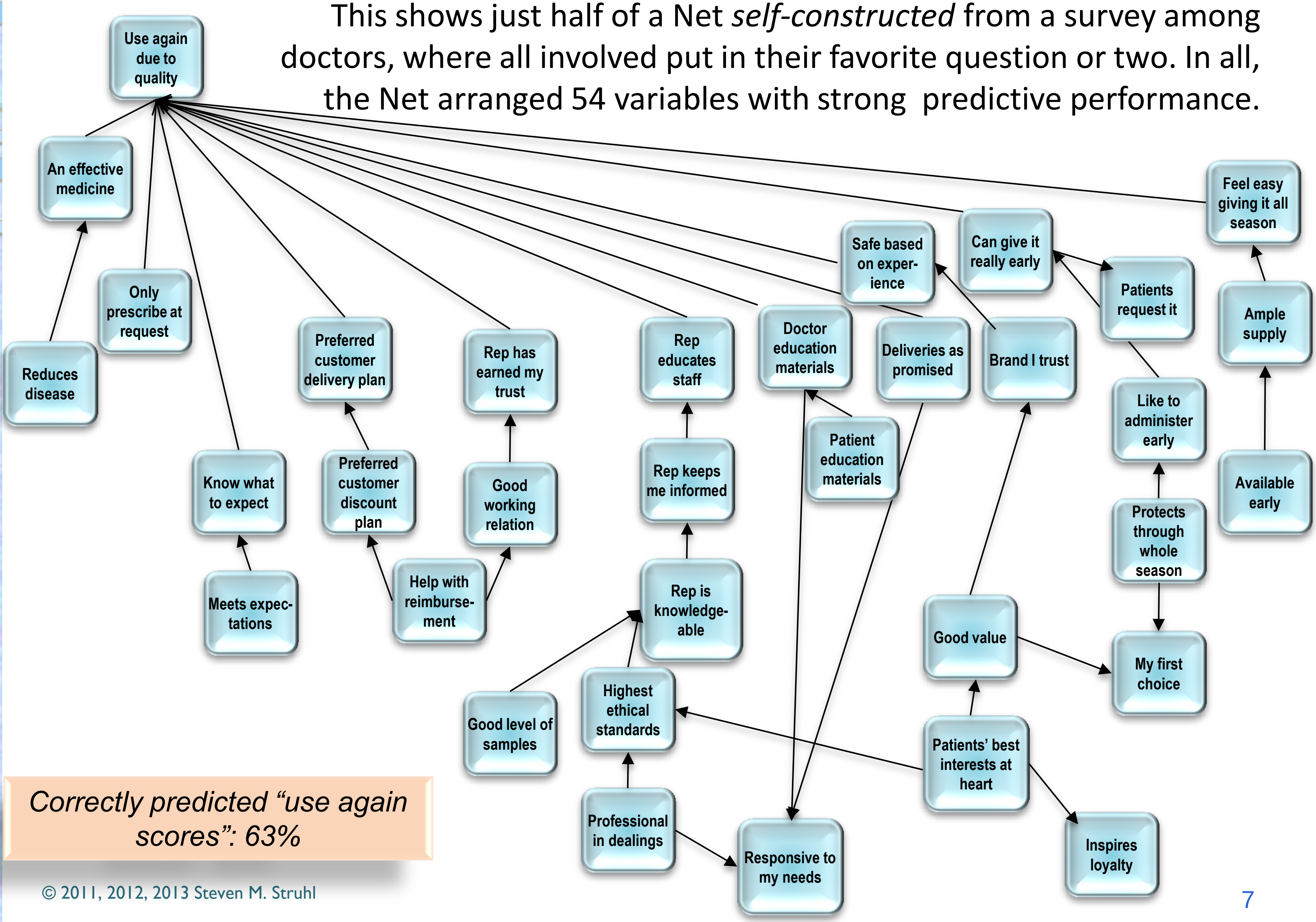
On to the story behind the data: Bayes Nets reveal how variables connect

- The Bayes Net importances arise from this model—these connections show how variables work together
- Note that "dedicated to my success" goes directly to share of new business and that all the other variables feed into it
  - That is, we can see that scores in this area depend on scores in others—in fact, *all scores are connected*
- These connections are critical because the way variables join with each other determines their effects on the target—and on each other
- These connections have a great deal of "face validity"—how they link makes sense, based on our understanding of this market
- Bayes Nets also can sort out much more complex sets of relationships—as we are about to see.



Bayes Nets can reveal structure from a messy survey

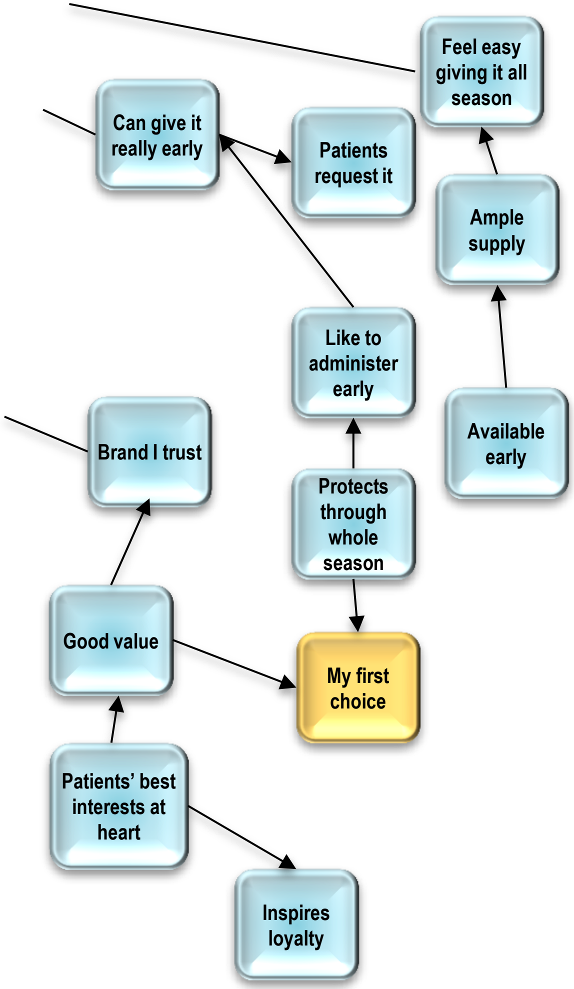
This shows just half of a Net *self-constructed* from a survey among doctors, where all involved put in their favorite question or two. In all, the Net arranged 54 variables with strong predictive performance.



Correctly predicted "use again scores": 63%

Bayes Nets: The connections make sense

- Let’s zero in on the variables closely related to “my first choice”—a focal point as connected by the Net itself
  - Note the closest connections are “good value” and “protects through the whole season”
    - These both seem to be relatively “hard edged” objective criteria
    - But then, “good value” is linked directly to “(has) patients’ best interests at heart”
    - So a drug is not as good a value unless the doctor also believes the maker cares about patients.
    - Influences on “good value” will move “my first choice”—the diagram shows that these are closely tied--so many areas matter
- Also, if we go back to the larger diagram, we learn more from seeing the variables *not* connected closely to “first choice”
    - These include the more relationship-oriented “responsive to my needs”—itself a focal point for other variables

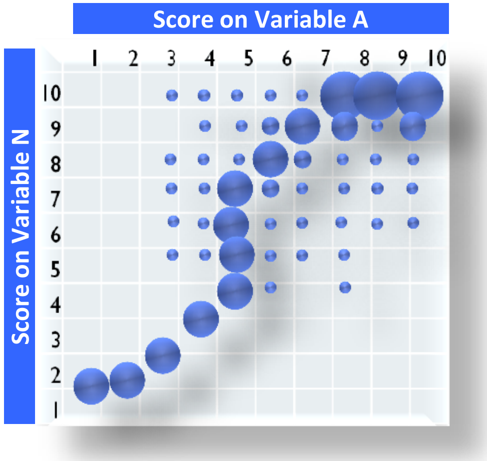


Why do Bayes Nets predict better?

- We can see how in this example, showing responses to two variables
- The boxes are counts of how responses align
  - So for instance, starting with the upper right corner:
    - 33 individuals gave a 10 rating to variable A and a 10 to variable N,
    - 44 gave a 9 to variable A and a 10 to variable N,
    - 45 gave an 8 to A and a 10 to N, and so on.

		Score on variable A										Totals
		1	2	3	4	5	6	7	8	9	10	
Score on variable N	10			1	1	1	6	4	45	44	33	135
	9				2	1	5	31	6	5	12	62
	8			1	3	4	29	11	5	7	8	68
	7			2	4	28	9	4	4	4	2	57
	6			3	4	21	7	5	3	2	2	47
	5			2	2	34	6	5	2			51
	4					32	5		1			38
	3				28							28
	2			31								31
	1	32	40									72
Totals		32	40	40	44	121	67	60	66	62	57	**589**

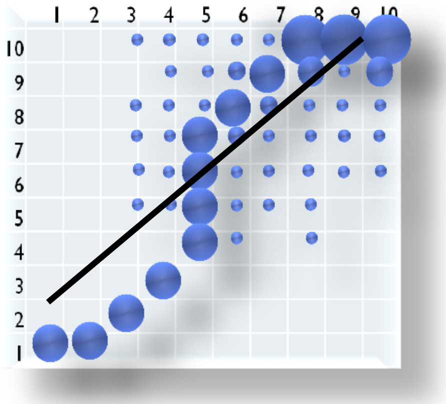
- This is the **"joint probability distribution"** of these two variables.



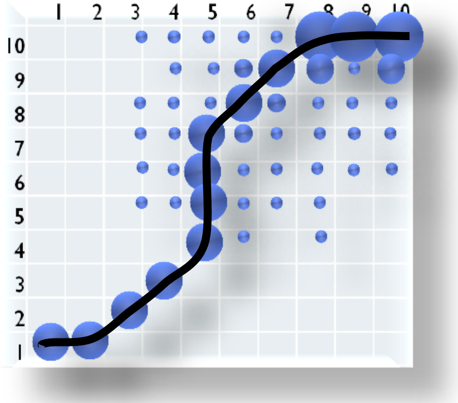
- We can use graphics to simplify this going forward:
  - Circles stand for how frequent responses are, with larger circles for more responses
- Let's see how Bayes Nets compare with regression-based models in predicting responses to variable N, using the responses to variable A . . .

Bayes Nets pick up the exact predictive pattern

- Regression and correlation-based models are “linear” and so only seek the best fitting straight line
  - They will not predict values well here
    - There are special “transformations” that can help with some kinds of non-linear data, but these are very hard for most audiences to follow

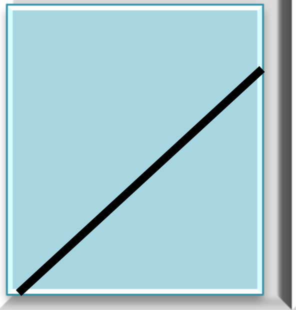
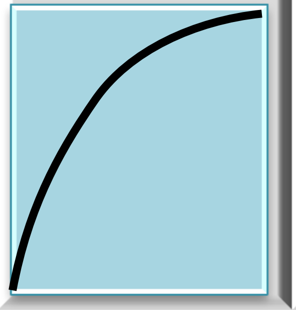
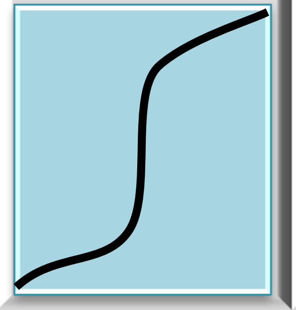
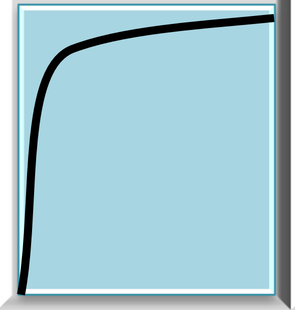
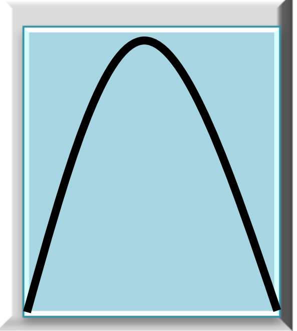


- Bayes Nets can model the relationship accurately, because they look at the entire distribution of each variable
  - That is, they use much more information than a summary of how closely two variables conform to a straight line



Overall comparison: What you can model successfully

Bayes Nets in fact capture any kind of regular relationship because they look at how variables fit together across all their values

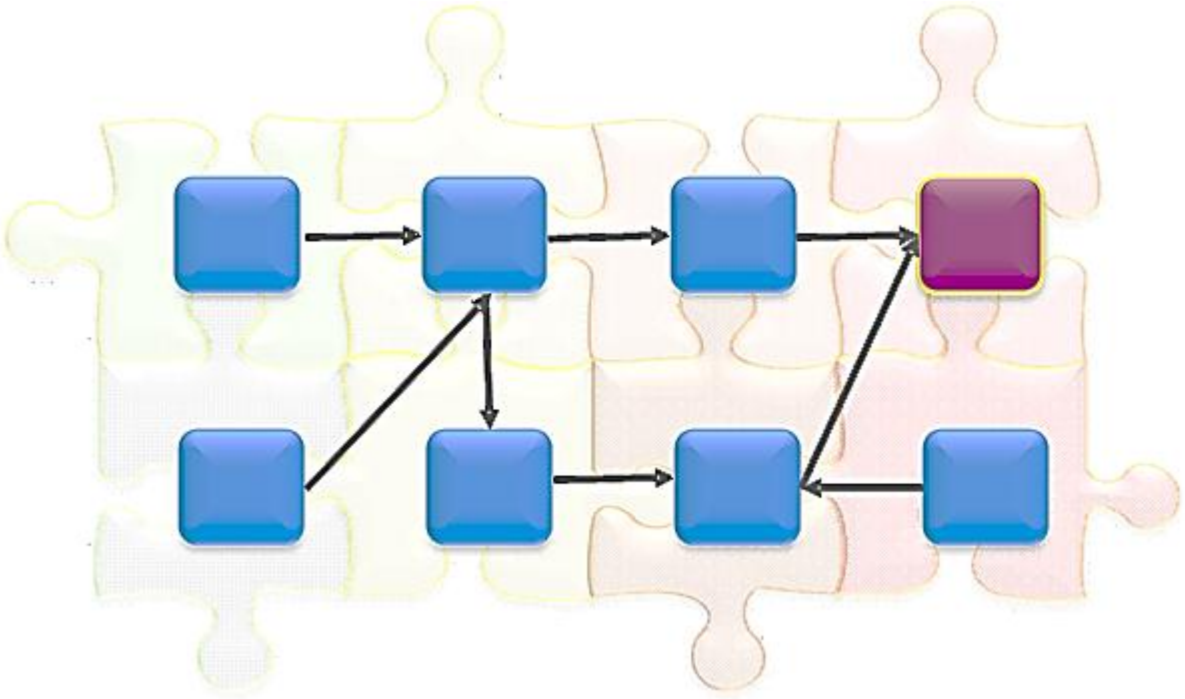
Type of relationship					
	Straight line	Decreasing returns curve	S-shaped (growth) curve	Rapid saturation curve	Best in middle ("just right") curve
Bayes Net	🟢	🟢	🟢	🟢	🟢
Regression (Includes PLS path models/ SEMs)	🟢	🟡	🟡	🔴	🔴

Legend	
🟢	Works well
🟡	Misses something**
🔴	Basically wrong

***Can be helped by doing special "transformations" before the analysis. These typically are not done, due to difficulties in explaining results.*

Bayes Nets find the connections in the data

- With Bayes Nets, the "data speaks"—revealing how variables link, and how they influence each other, as well influence the target variable
- Sophisticated modeling procedures will try hundreds or thousands of arrangements, fitting the variables together like pieces of a puzzle until reaching a best overall model



Original illustration by L. da Vinci

Basic background on Bayes Nets

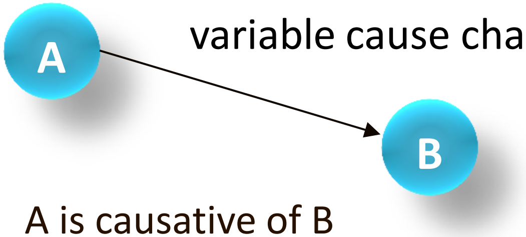
More of the story behind the results

Networks show directions and connections

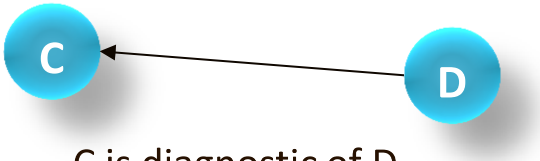
- A network is a “directed acyclic diagram” (or DAG): there must be directions between variables
- The variable at the start of an arrow is called a “parent”
  - The variable at the end is called a “child” of the parent
  - The parent node leads to (and in the right circumstances *can cause*) the child node
    - Arrows can lead to or from a dependent variable
    - Children can have several parents and parents can have several children

For once, the terminology is warm and fuzzy

- Many arrows work as well in either direction
  - They are “equivalencies” and we must decide how to orient them
    - Most arrows we encounter with survey data are equivalencies
    - Whatever the directions, **influence flows both ways**—changes in one variable cause changes in the other.



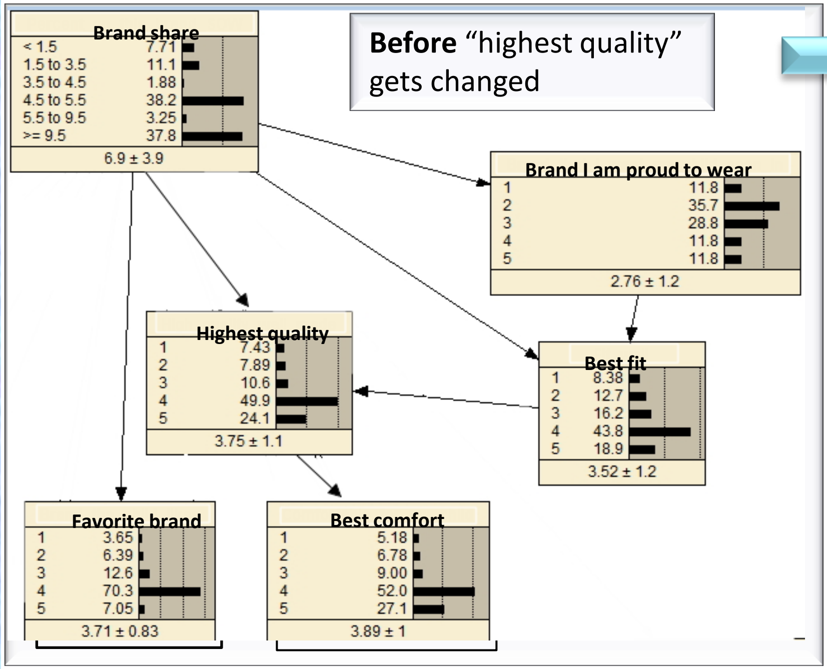
A is causative of B
or leads to B



C is diagnostic of D
or explains D

Nets measure effects more accurately

- Nets measure more accurately because *all variables connect*.
  - Changes *propagate* throughout a network—when one variable changes all will change. So effects take into account all other variables—a huge advance over regression—where "all other things are assumed to remain the same"
- Here is a Net showing the **distribution** of each variable—the percent of answers falling into each scoring range
  - From left to right, we have changed one variable (shown by the box's gray color—we boosted top scores in "highest quality" by 10%).
  - When we do ***everything else changes***—not just the target variable, share.



Before "highest quality" gets changed

Brand share	
< 1.5	7.71
1.5 to 3.5	11.1
3.5 to 4.5	1.88
4.5 to 5.5	38.2
5.5 to 9.5	3.25
>= 9.5	37.8
6.9 ± 3.9	

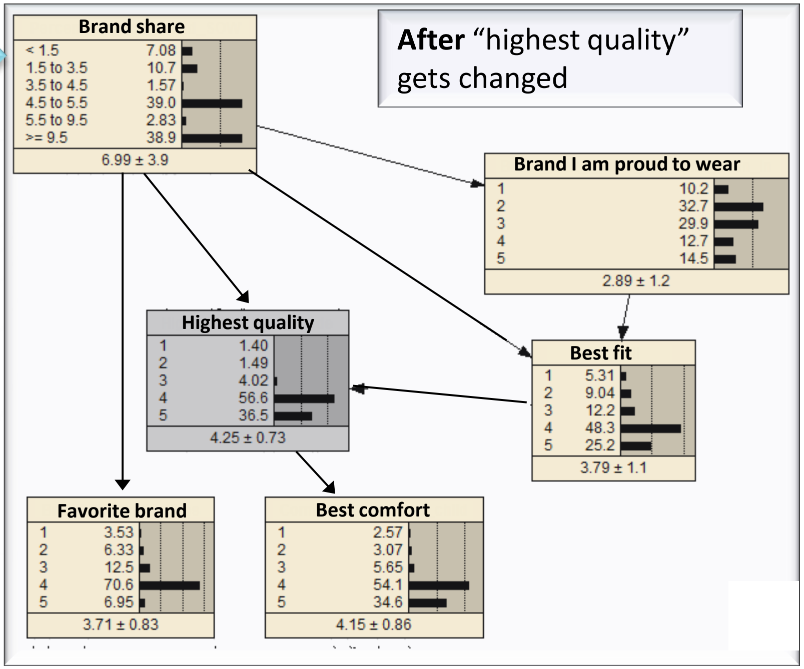
Brand I am proud to wear	
1	11.8
2	35.7
3	28.8
4	11.8
5	11.8
2.76 ± 1.2	

Highest quality	
1	7.43
2	7.89
3	10.6
4	49.9
5	24.1
3.75 ± 1.1	

Best fit	
1	8.38
2	12.7
3	16.2
4	43.8
5	18.9
3.52 ± 1.2	

Favorite brand	
1	3.65
2	6.39
3	12.6
4	70.3
5	7.05
3.71 ± 0.83	

Best comfort	
1	5.18
2	6.78
3	9.00
4	52.0
5	27.1
3.89 ± 1	



After "highest quality" gets changed

Brand share	
< 1.5	7.08
1.5 to 3.5	10.7
3.5 to 4.5	1.57
4.5 to 5.5	39.0
5.5 to 9.5	2.83
>= 9.5	38.9
6.99 ± 3.9	

Brand I am proud to wear	
1	10.2
2	32.7
3	29.9
4	12.7
5	14.5
2.89 ± 1.2	

Highest quality	
1	1.40
2	1.49
3	4.02
4	56.6
5	36.5
4.25 ± 0.73	

Best fit	
1	5.31
2	9.04
3	12.2
4	48.3
5	25.2
3.79 ± 1.1	

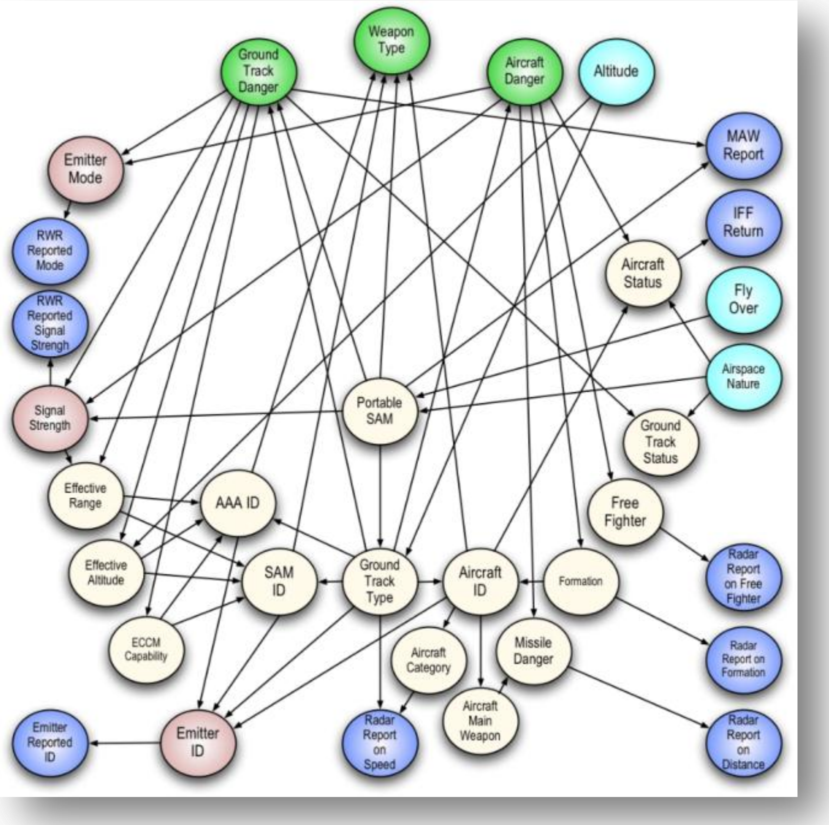
Favorite brand	
1	3.53
2	6.33
3	12.5
4	70.6
5	6.95
3.71 ± 0.83	

Best comfort	
1	2.57
2	3.07
3	5.65
4	54.1
5	34.6
4.15 ± 0.86	

Note that even "favorite brand" changes a little, although it is not directly connected to "highest quality".

Networks self-construct trying many alternatives

- The network typically is the result of countless attempts by the program to fit together the data—seeing how variables best work to predict the target variable
  - At least a dozen methods for growing networks are readily available
    - Testing more than one can be required to get the best Net
- As well as the best methods do, though, we must be the ultimate arbiter of what makes sense
  - We always have the ability to tune, tweak and test the network
- A well constructed network typically will be very powerful and will predict remarkably well



Networks are battle-tested, quite literally (at the center is a SAM or "surface to air missile"). This network weighs all the factors deciding whether to launch it.

Nets can perform many other useful tasks

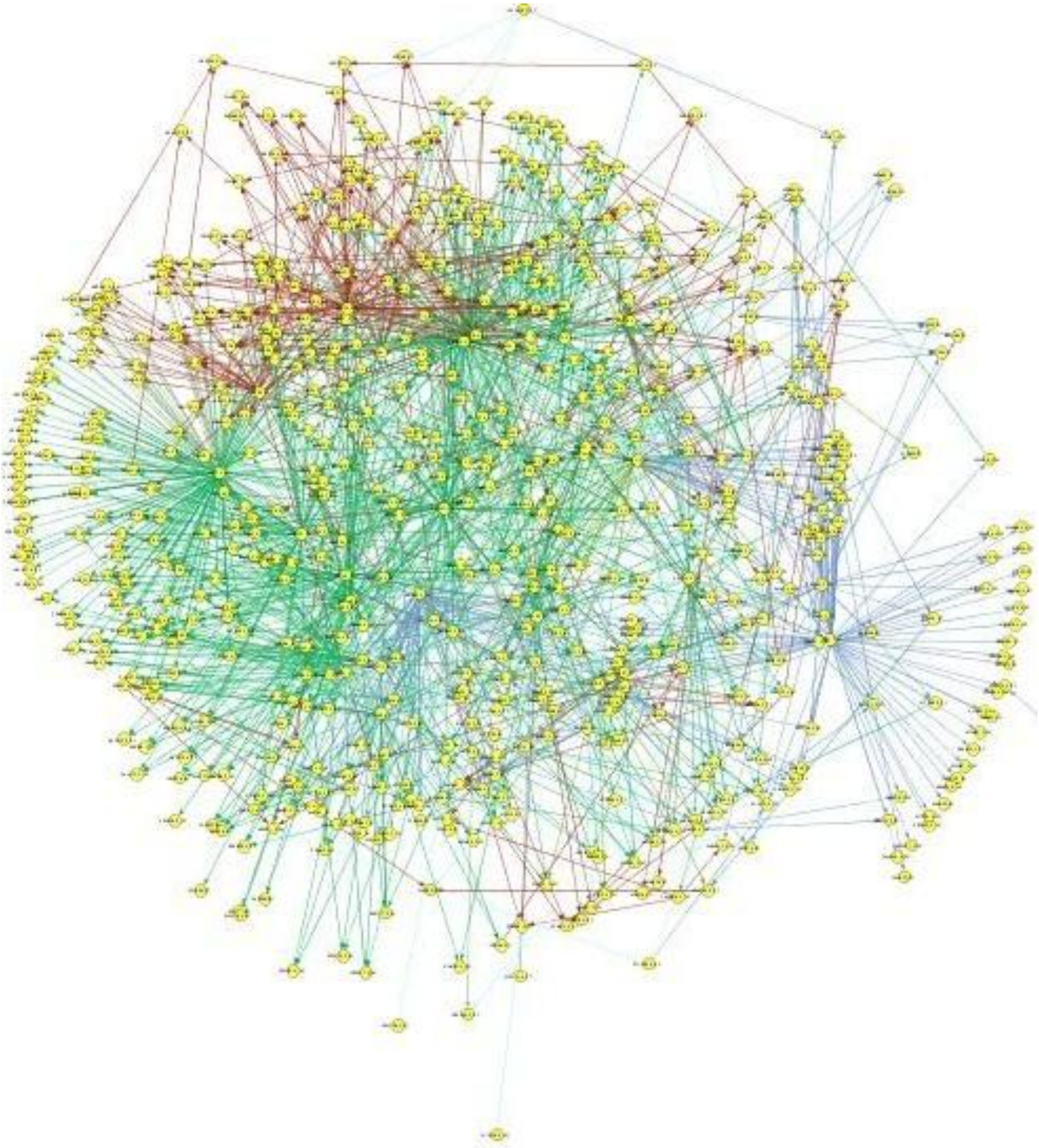
- Developing models of cause and effect (where the data supports this)
- Incorporating expert judgment into models
- Determining far more accurately than cross-tabulations how sparsely represented groups are likely to respond

- Induction by automatic learning

  - Data mining
  - Web analysis
  - Text analysis
  - Anywhere variables need to be tested or winnowed
- Many other applications from brainstorming to the most sophisticated modeling
- Uses are still expanding.

Bayes Nets work in serious applications

- Most of us will never see a Net like this in action—and, staggering as it is, it works
- It diagnoses the causes of B cell chronic lymphocytic leukemia (B-CLL)
- Here and elsewhere, Nets are trusted as the best analytical method in actual life-and-death applications, so we definitely can rely on the them with our data.



There is a lot more, so please ask questions

Hello? Help line?
I appear to have misplaced my sense of fun.

We can answer many of them

Dr. Steven Struhl

smstruhl@gmail.com