Pseudodiagnosticity misdiagnosed?

Rachael Bond

Tom Ormerod

Linden Ball

July 26th 2013,

London Reasoning Workshop

LANCASTER UNIVERSITY

Overview

1. Pseudodiagnosticity: the phenomenon
2. Introducing entropy
3. The "Maximum Entropy Reduction" model
4. A computational exploration of Pseudodiagnosticity
5. Empirical testing
6. Conclusions

What is Pseudodiagnosticity?

· Pseudodiagnosticity was first identified by Michael Doherty et al in 1979.

· What they noticed was that their participants seemed to select worthless information when given decision making tasks.

· By "worthless" Doherty was referring to information which, in his opinion, could not help the participants complete their tasks.

· Because the participants thought they were choosing useful information, Doherty decided to call this phenomenon "pseudodiagnosticity".

· In other words, "pseudodiagnosticity" refers to choosing information to help solve a problem that appears to be useful when really it isn't.

Reference: Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticitgy.
Acta Psychologica, 43(2), 111–112. doi: 10.1016/0001-6918(79)90017-9

Why was the information "worthless"?

1. The exercises that Doherty set were based on Bayes' theorem.

2. Bayes' theorem uses conditional probabilities to calculate the likelihood ratios for different hypotheses. The result is called the "Bayes' factor".

3. Using Bayes' theorem requires information for all the options (hypotheses) and Doherty's participants tended to choose information relating to just one hypothesis.

4. Bayes' formula is given as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

How does this work?

1. Doherty et al (1979) asked their participants to imagine that they were an under-sea explorer who had just discovered an ancient pot.

2. They were told that the pot could have come from either of two nearby islands and that it was their job to find out which.

3. Some information was then provided to help the participants diagnose the pot's origins...

Reference: Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticitgy. *Acta Psychologica, 43(2)*, 111–112. doi: 10.1016/0001-6918(79)90017-9

The Pseudodiagnosticity Information Matrix

The participants were told that the pot they had found had curved handles and was made from a smooth clay.

They were then given an information matrix which looked like this:

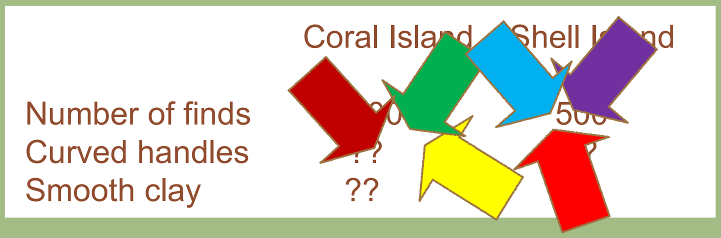
	Coral Island	Shell Island
Number of finds	5000	500
Curved handles	??	??
Smooth clay	??	??

and told that they could reveal two of the hidden pieces of information to help them make a decision.

What Doherty expected...

was for participants to choose data pairs.

What Doherty actually got was...



a bit of a mess!

What was Doherty et al's conclusion?

"Pseudodiagnosticity is clearly dysfunctional"

Doherty et al, 1979 p. 121

Reference: Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticitgy.
Acta Psychologica, 43(2), 111–112. doi: 10.1016/0001-6918(79)90017-9

Is Pseudodiagnosticity *really* dysfunctional?

The Pseduodianosticity paradigm is, nowadays, most commonly presented with one piece of diagnostic information (the “anchor information”). So, adopting this format, imagine if the participants had been given this (extreme) information:

	Coral Island	Shell Island
Number of finds	900	100
Curved handles	90%	??
Smooth clay	??	??

· The only situation in which Coral island would not necessarily be the correct answer is if the figure for "Smooth Clay|Coral Island" results in an overall find rate for this type of pot of less than 100.

· Selecting the information for Shell Island does not help since it can not refute the hypothesis that the pot came from Coral island.

· The information for Shell Island would only have been useful if the participants had been asked to estimate the likelihood for each hypothesis rather than just to make a decision.

But what if Doherty *had* asked for likelihoods?

Even if the participants had chosen paired information, like this:

	Coral Island	Shell Island
Number of finds	900	100
Curved handles	90%	80%
Smooth clay	??	??

· The question indicates that an answer must take into account that the pot has curved handles and is made from smooth clay.

· An answer can only be given if estimates are made for missing information.

· Although there is no effect on the calculated Bayes' factor if both estimates are the same (they cancel out), their logical existence must be acknowledged.

· There is no mathematical reason why these estimates have to be applied to the same diagnostic feature ("curved handles", "smooth clay" etc.). In the long run any inaccuracies will even out.

Introducing Entropy (1)

In our example, the attraction of selecting "Smooth Clay|Coral Island" stems from the relative magnitude of the Coral Island figures compared to those for Shell Island.

	Coral Island	Shell Island
Number of finds	900	100
Curved handles	90%	??
Smooth clay	??	??

· To put it another way, there is a greater variation in the potential values for Coral Island than Shell Island (1-810 vs. 1-100).

· This variation can also be thought of as "disorder", for which a value may be calculated using Shannon's (1948) entropy formula (for discrete variables):

$$H(X) = -\sum_{i=1}^{n} p(x_i) log_b p(x_i)$$

Reference: Shannon, C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379–423. Retrieved from http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf

Introducing Entropy (2)

· A "Maximum-Entropy estimate" (MaxEnt) may be used for the missing values.

· MaxEnt may be defined as:
the *“least biased estimate possible on the given information; i.e., it is maximally noncommittal with regard to missing information”* (Jaynes, 1957, p. 620).

· The least biased estimate is provided by the standard deviation curve.
(Park & Brera, 2009)

· The substitution value may be calculated as the area under the curve as a percentage of its universe. Normally this will be 0.5.

· This figure is not necessarily the same as the "value of greatest indifference" as suggested by Crupi, Tentori, and Lombardi (2009).

References: Crupi, V., Tentori, K., & Lombardi, L. (2009). Pseudodiagnosticity revisited. *Psychological Review, 116 (4),* 971–985. doi: 10.1037/a0017050
Jaynes, E.T. (1957). Information theory and statistical mechanics. *Physical Review, 106 (4),* 620–630. doi: 10.1103/PhysRev.106.620
Park, S.Y., & Bera, A.K. (2009). Maximum entropy autoregressive conditional heteroskedasticity model. *Journal of Econometrics, 150 (2)*, 219–230. doi: 10.1016/j.jeconom.2008.12.014

Introducing Entropy (3)

Using MaxEnt it is now possible to calculate the entropy, *H*, for each hypothesis:

	Coral Island	Shell Island
Number of finds	900	100
Curved handles	0.9	0.5
Smooth clay	0.5	0.5

$$H(Coral) = -(\frac{1}{900 \times 0.9 \times 0.5} \times log_2 \frac{1}{900 \times 0.9 \times 0.5}) \times 900 \times 0.9 \times 0.5 = log_2(405) = 8.66\, bits$$

$$H(Shell) = -(\frac{1}{100 \times 0.5 \times 0.5} \times log_2 \frac{1}{100 \times 0.5 \times 0.5}) \times 100 \times 0.5 \times 0.5 = log_2(25) = 4.64\, bits$$

· Since, for comparison purposes, the use of logs cancels out, a selection strategy based on entropy reduction becomes the simple heuristic of choosing the hypothesis with the largest value.

· We have termed this the "Maximum Entropy Reduction" strategy (MER).

How does MER differ from Information Gain?

· The Maximum Entropy Reduction strategy simplifies the Information Gain model.

· Where the Information Gain model suggests that:
"the difference between the [post selection] uncertainty and the prior uncertainty... indicates the gain in information provided by a piece of data" ...
(Oaksford & Chater, 2007, p. 170)

... MER is only concerned with the absolute levels of entropy as percieved by the decision maker.

Reference: Oaksford, M., & Chater, N. (2007). *Bayseian rationality. The probabilistic approach to human reasoning*. Oxford, England: Oxford University Press.

Modelling Pseudodiagnosticity

To test the MER strategy a computer model calculated the results for the four different selection strategies of "pair", "column", "diagonal" and "MER".

	H_1	H_2
Prior Information	**10-90**	**90-10**
D_1	**"anchor info"**	**"pair"**
D_2	**"column"**	**"diagonal"**

· Prior frequency ranges of 10-90 were used to generate discrete posterior frequencies.

· The Bayes' factor for every combination was calculated and the result was compared to the answers given by each selection strategy.

· This comparison was made for categorical decisions as well as for an estimation of likelihoods (the Bayes' factor).

LANCASTER UNIVERSITY

Modelling Pseudodiagnosticity: Results (1)

These are the success rates, by prior information, for each selection strategy when a categorical decision is made.

Priors	# combinations	H_2/D_1 Correct		H_2/D_2 Correct		H_1/D_2 Correct		MER Correct*	
		Total	%	Total	%	Total	%	Total	%
10:90	810000	726786	89.7%	726786	89.7%	706793	87.3%	726786	89.7%
20:80	2560000	2150018	84.0%	2150018	84.0%	1997181	78.0%	2150018	84.0%
30:70	4410000	3494897	79.2%	3494897	79.2%	3206190	72.7%	3494897	79.2%
40:60	5760000	4355968	75.6%	4355968	75.6%	4140276	71.9%	4288954	76.4%
50:50	6250000	4622650	74.0%	4622650	74.0%	4638956	74.2%	4716595	77.0%
60:40	5760000	4355968	75.6%	4355968	75.6%	4486966	77.9%	4514121	79.7%
70:30	4410000	3494897	79.2%	3494897	79.2%	3638640	82.5%	3632843	83.6%
80:20	2560000	2150018	84.0%	2150018	84.0%	2235001	87.3%	2222006	87.9%
90:10	810000	726786	89.7%	726786	89.7%	750580	92.7%	744049	92.9%
	33330000	26077988	78.2%	26077988	78.2%	25800583	77.4%	26490269	80.6%

Note. *The percentages given for the Maximum Entropy Reduction strategy are based on the number of occasions in which there was a difference between the perceived entropy values of H_1 and H_2

Modelling Pseudodiagnosticity: Results (2)

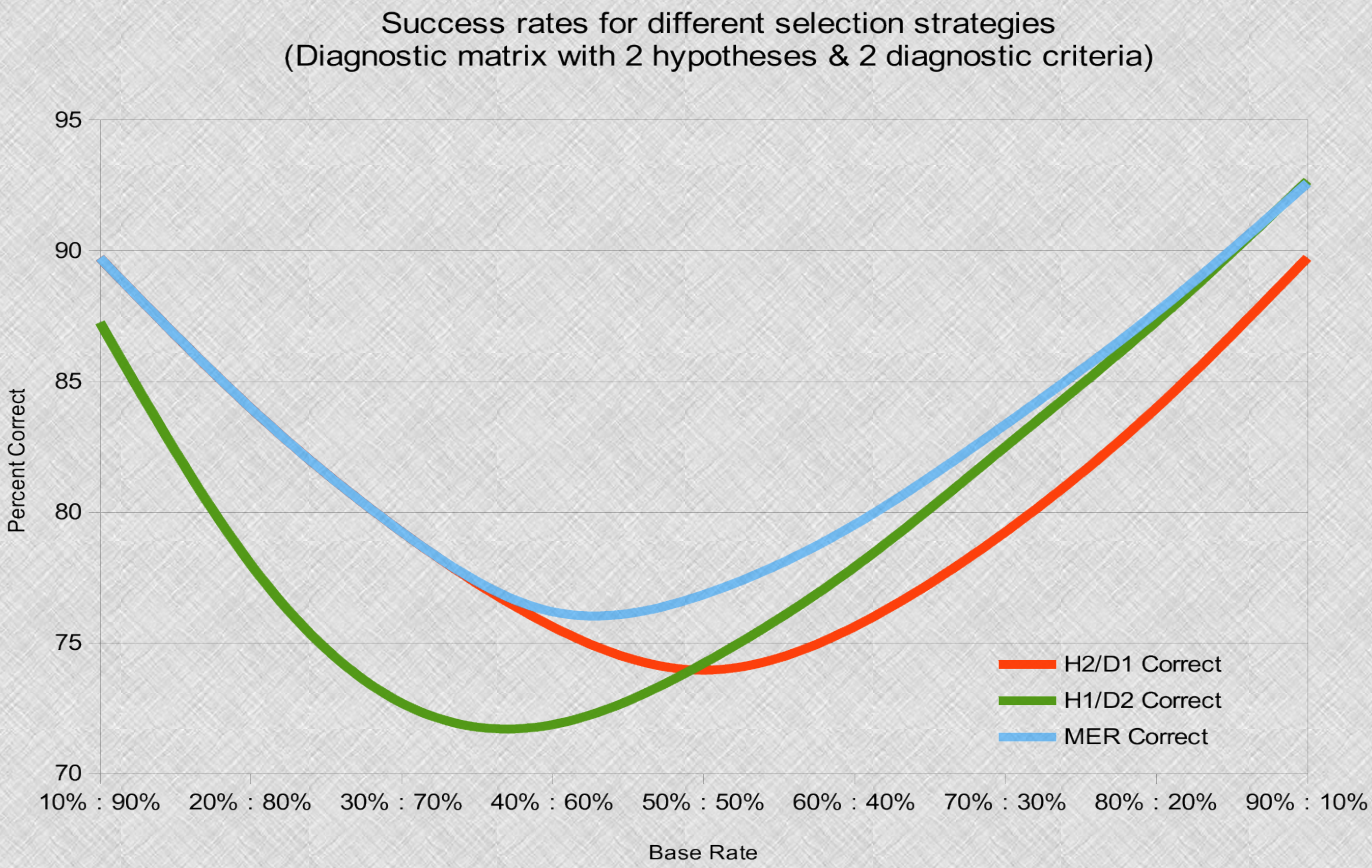
· As one would expect there was no difference between selecting "pair" or "diagonal".

· This means that the choice for the decision maker lies solely between the different hypotheses - the choice of differentiating characteristic is irrelevant.

· For even priors and above, selecting "column" outperforms selecting "pair".

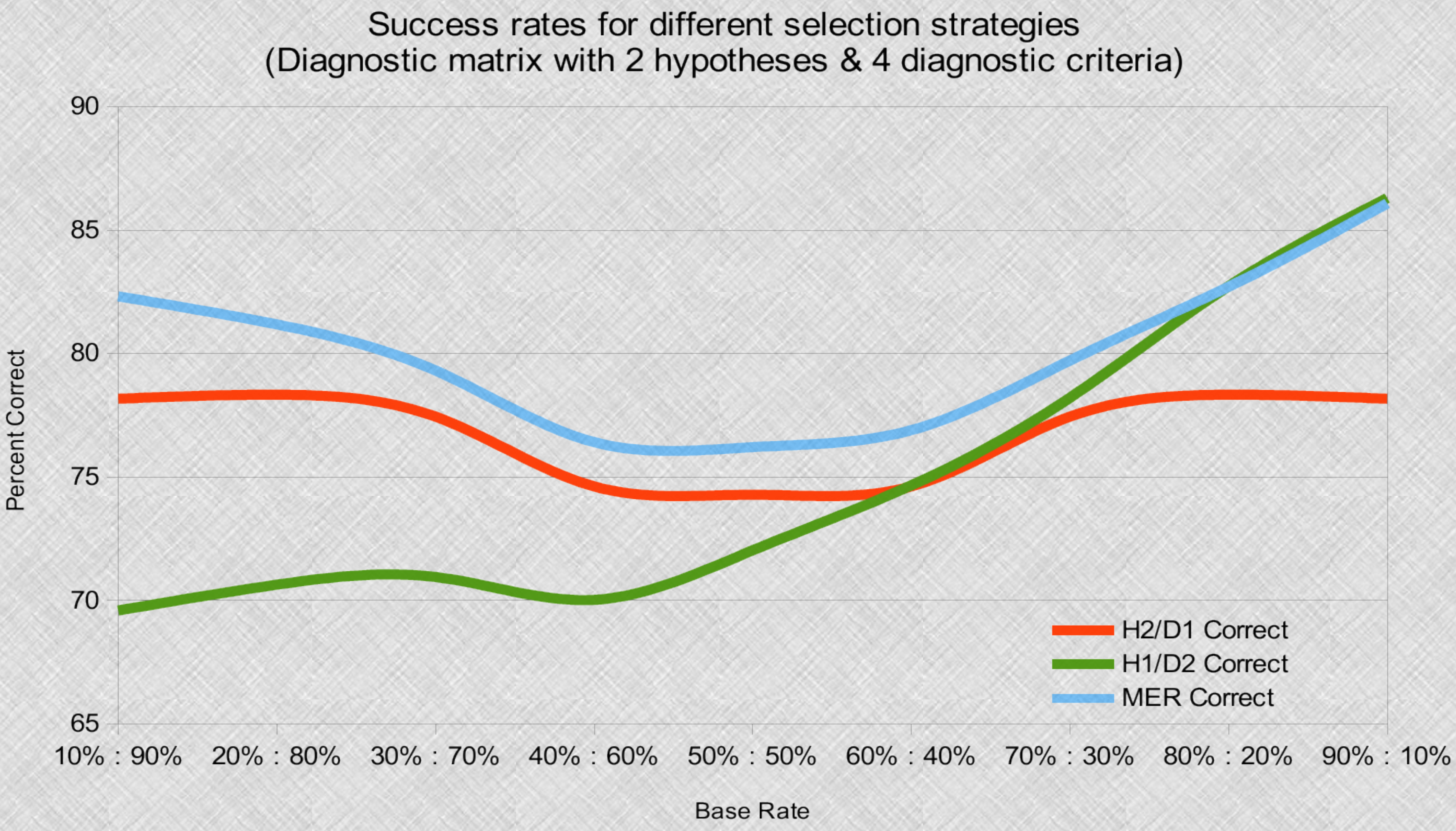
· The best performing strategy is MER, which is seemingly able to correctly select between "pair" and "diagonal", as shown graphically in the following slide.

Modelling Pseudodiagnosticity: Results (3)



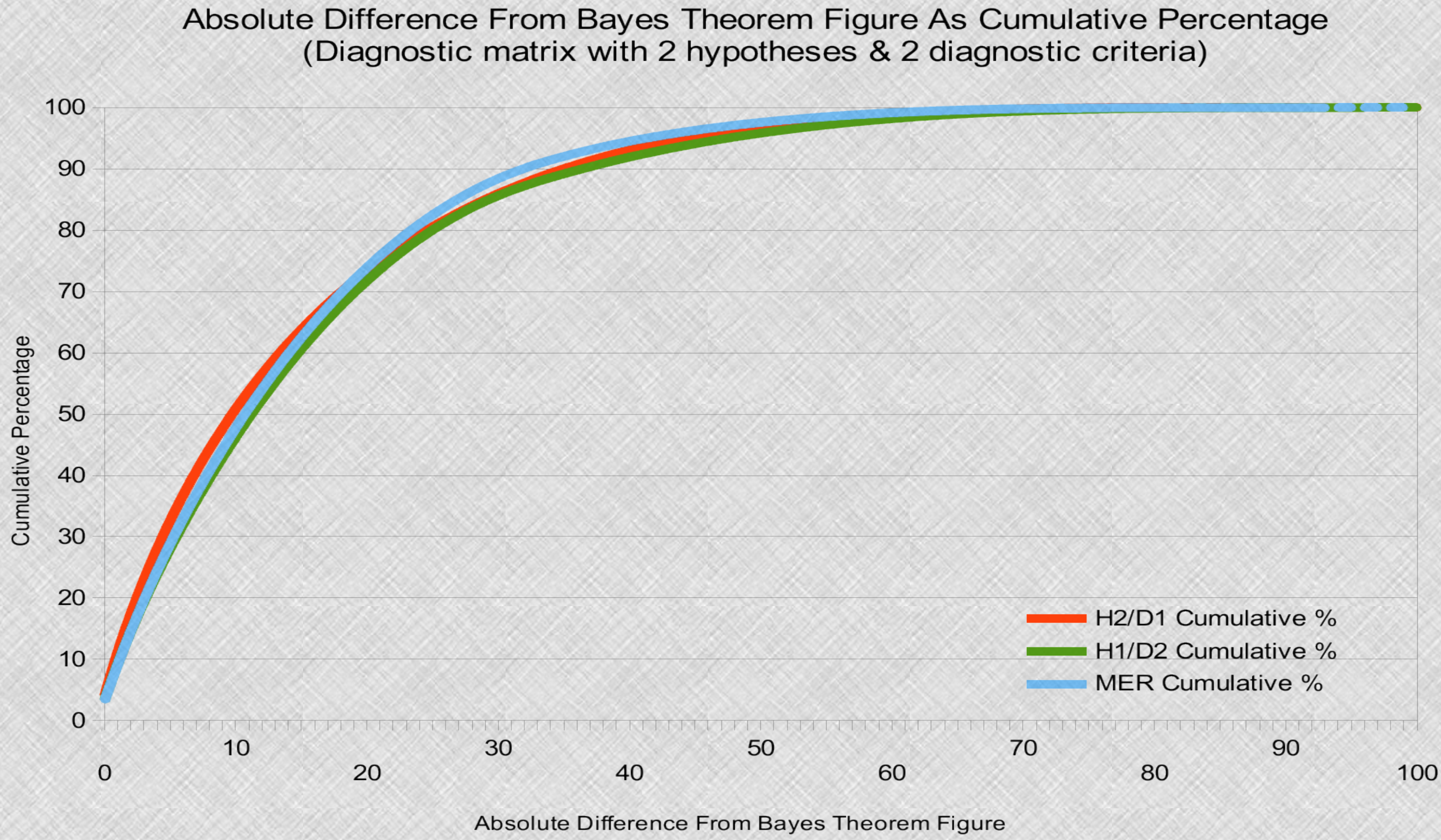
Modelling Pseudodiagnosticity: Results (4)

- If the diagnostic matrix is extended to include four differentiations & three data selections, MER can still be seen to be the more powerful selection strategy.



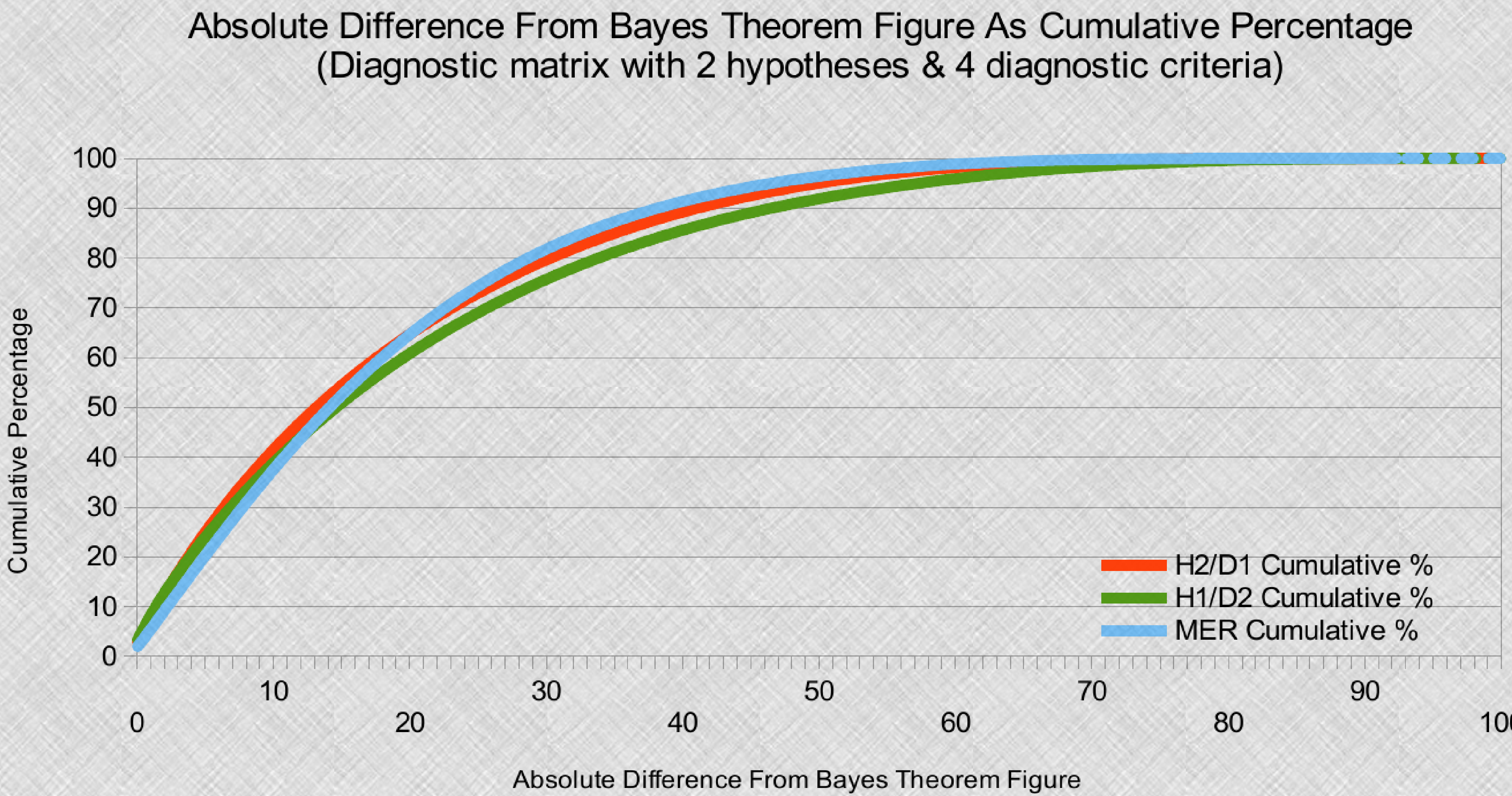
**Note*. For computational simplicity the event space was sampled in units of three (16,710,298,374 trials).

Modelling Pseudodiagnosticity: Results (5)



· For estimating the Bayes' figure, selecting "pair" (H_2/D_1) is most likely produce an estimate within +/- 10%. However, it is MER which produces the lowest overall error.

Modelling Pseudodiagnosticity: Results (6)



- Once again the same pattern emerges if the diagnostic matrix is extended to include four differentiations and three data selections.

**Note*. For computational simplicity the event space was sampled in units of three (16,710,298,374 trials).

Modelling Pseudodiagnosticity: Implications

· The MER strategy outperforms the selection of paired data when making a categorical choice between hypotheses.

· The average absolute estimation error from the Bayes' factor is lower for the MER strategy than the selection of paired data.

· For these reasons the normativity of MER for making both categorical decisions and estimating likelihood ratios, over the entire prior and posterior sample space, is asserted.

Putting the theory to the test

The discussion leads to four research questions:

1. Is there evidence that people follow a MER strategy when searching for information?

2. Is there evidence that people will adopt different search strategies when making categorical decisions rather than estimating likelihood ratios?

3. Is there evidence that people will adopt different search strategies when answering subjective, rather than diagnostic, questions?

4. Is there evidence that people adopt different search strategies according to pseudodiagnostic matrix size and complexity?

The experiment (1)

· A publicly accessible research website was coded.

· The participants were presented with a series of six decision making tasks comprising of three diagnostic (objective) questions and three subjective questions to which they were requested to provide either a categorical answer or likelihood estimates.

· Each task was constructed using either two hypotheses with two diagnostic criteria, two hypotheses with four diagnostic criteria or three hypotheses with four diagnostic criteria.

· For each exercise one piece of anchor information from the first diagnostic criterion was provided.

The experiment (2)

· All aspects of the research were either randomly generated or counterbalanced: the prior and posterior data were randomly generated; the six question texts were randomly allocated to one of the three diagnostic matrix sizes and to one of the two response styles; the presentation order of the questions was randomized; and the position of the anchor information was counterbalanced.

· The use of an online experimental format helped establish ecological validity, allowed a wider participant demographic than is often used in psychology research, and ensured that the results were free from any inadvertent researcher intervention.

· The results of the first 150 participants who successfully completed all the exercises were accepted. There were no exclusions and there was no inducement to participate.

· The participants (*N*=150) were 35% male and 63% female (2% undeclared) with an even distribution of age from 18 to 60 years.

Example exercise (3 hypotheses & 4 diagnostic criteria)

Your friend has a car she bought a couple of years ago. It's either made by "Solus", "Trisor" or "Tomcat", but you can't remember which. You do, however, remember that her car:

1: does over 25 miles per gallon

2: has not had any major mechanical problems in the two years she's owned it

3: has four wheel drive

4: is a hatchback

Below is a table giving some information about the cars made by the three companies. Using this information you must decide whether your friend's car was made by "Solus", "Trisor" or "Tomcat". To help you make your decision you may select ***five*** further pieces of information from the table by clicking on the red squares.

	Solus	Trisor	Tomcat
The total number of cars sold by each company	1000	5000	4000
The percentage of each make which does over 25 mpg		30%	
The percentage of each make with no major mechanical problems in the last two years			
The percentage of each make that has four wheel drive			
The percentage of each make that is a hatchback			

Which company made your friend's car?

Solus | Trisor | Tomcat

>> I am happy with my decision >>

Example exercise (2 hypotheses & 2 diagnostic criteria)

You are thinking about booking a holiday, but can't decide where to go. You make a list of your requirements:

1: there should be good beaches

2: there should be a great nightlife

and look on tripadvisor.com to see where people recommend going. Below is a table giving some information about two destinations, "Puerto Blanco" and "Villa Negra". Using this information you must estimate the likelihood of each destination being suitable for your holiday. To help, you may select ***one*** further piece of information from the table by clicking on one of the three red squares.

	Puerto Blanco	Villa Negra
The number of people who have been	700	300
The percentage of each destination's beaches which are rated as good		50%
The percentage of people who rate each destination's nightlife as great		

What are the chances that each destination will be suitable for your holiday?

(Note: the combined chances must equal 100%)

Puerto Blanco: [ ] Villa Negra: [ ]

>> I am happy with my decisions >>

The experiment: Results (1)

The gross cell selection patterns (%) for the matrices with 2 hypotheses and 2 diagnostic criteria were:

Diagnostic questions*		Subjective questions*		Categorical decisions*		Likelihood estimates*		Anchor H_1/D_1		Anchor H_2/D_1	
-	36	-	36	-	36	-	36	-	35	37	-
27	37	22	42	21	43	28	36	21	43	35	28

Note. *Where the anchor information was presented to the participants in H_2/D_1 the results have been reflected; i.e., inverted.

· Compared to an even distribution, χ^2 tests give $\chi^2(2,N=100)=7.59, p=0.023$ for questions requiring a categorical choice decision and $\chi^2(2,N=100)=1.28, p=0.527$ for those requiring likelihood estimates.

· When the anchor information is given in H_1/D_1, $\chi^2(2,N=100)=7.44, p=0.024$ when compared to an even distribution, and $\chi^2(2,N=100)=1.34, p=0.512$ when the anchor information is given in H_2/D_1.

· For diagnostic questions $\chi^2(2,N=100)=1.82, p=0.403$ and $\chi^2(2,N=100)=6.32, p=0.042$ for subjective questions.

The experiment: Results (2)

Comparison of cell selections against the MER strategy:

· Overall there was no goodness of fit between the MER strategy and cell selections for the matrices with two hypotheses and two diagnostic criteria.

· However, where the anchor information was given in H_1/D_1 the overall goodness of fit was significant, $\chi^2(1,N=147)=4.55, p=0.0329$, but became highly insignificant when the information was presented in H_2/D_1 , $\chi^2(1,N=150)=0.12, p=0.729$.

· This is consistent with the gross selection patterns given earlier.

· For the matrices with three hypotheses and four diagnostic criteria an overall highly significant correlation is achieved for the first four cell selections with the results ranging from $\chi^2(1,N=291)=15.85, p<0.0001$ to $\chi^2(1,N=288)=23.44, p<0.0001$.
The fifth data selection point fails to reach significance.

· The results for subjective questions are also highly significant for the first four selections, whereas the diagnostic questions reveal some variability.

· Both response types of categorical decision and likelihood estimation are either significant or highly significant for the first four selections.

The experiment: Results (3)

· The fifth cell selection in the matrices with three hypotheses and four diagnostic criteria show a reversal of selection strategy. In particular the results for the subjective questions with a likelihood estimation show that only 17 selections were correctly predicted by MER as opposed to an expected figure of 22.2; $\chi^2(1,N=67)=1.49, p=0.22$.

· The results for the matrices with two hypotheses and four diagnostic criteria fell between the others with a trend towards an overall significance for goodness of fit for the first two cell selections, $\chi^2(1,N=298)=3.48, p=0.0621$ & $\chi^2(1,N=294)=3.71, p=0.0541$.
However, as with the three hypotheses matrix, there was no fit for the final selection.

Discussion

· The standard pseudodiagnosticity paradigm contains two of the features which demonstrably produce biased selection patterns: categorical decisions and the placing of anchor information in H_1/D_1.

· Can an apparent effect that is largely negated simply by reflecting the presentation of data still be considered to provide adequate support for the very important claims which it underpins?

· The predictive value of MER increases with matrix complexity.

· This could be because participant engagement increases with task complexity...

... Or it could be that as the degrees of freedom available to the participants reduce, there is an increased tendency to verify the estimated values of the alternative hypothesis.

· This would explain the reversal of strategy for the final cell selection shown in both the two and three hypotheses matrices with four diagnostic criteria. For the two hypotheses and two diagnostic criteria matrix, the apparent randomisation of selections would reflect the utility the individual participant places on gaining further knowledge against the verification of their assumptions about missing data values when there is only one degree of freedom available to them.

Conclusions

· The assertion of the Pseudodiagnosticity paradigm as to the normativity of the selection of data pairs has been shown to be incorrect.

· Instead, the "Maximum Entropy Reduction" strategy has been shown to be normative for both categorical decisions and the estimation of likelihood ratios.

· Where a matrix is sufficiently complex, there is a high correlation between participant information selection patterns and those predicted by MER.

· The reversal of strategy identified on the final cell selection is a new phenomenon. It is possible that this results from the reduction in available degrees of freedom, however further research is required to fully explain the implications of this finding.

The Researchers

Rachael Bond

email: hello@rachael.be
twitter: @rachael_bond
web: www.rachaelbond.com

Professor Tom Ormerod

email: t.ormerod@lancaster.ac.uk

Professor Linden Ball

email: LBall@uclan.ac.uk