Information Relevance in Pseudodiagnostic Reasoning

Frédéric Vallée-Tourangeau and Gaëlle Villejoubert
Psychology Research Unit, Kingston University
Kingston-upon-Thames UNITED KINGDOM KT1 2EE
f.vallee-tourangeau/g.villejoubert@kingston.ac.uk

Abstract

When faced with two competing hypotheses, people sometime prefer to look at multiple sources of information in support of one hypothesis rather than to establish the diagnostic value of a single piece of information for the two hypotheses. This is termed pseudodiagnostic reasoning, and is understood to reflect a pervasive confirmation bias. Past research suggests that diagnostic reasoning may be more easily fostered when participants seek data to help in the selection of one of two competing courses of action as opposed to situations where they seek data to help inferring which of two competing hypotheses is true. In the experiment reported here, we provide the first empirical evidence demonstrating that the facilitating effect observed in action problems is driven by considerations of information relevance, reasoners' motivations and the numerical value of the first piece of information presented. The discussion of these findings focuses on implications for the ability to engage in diagnostic hypothesis-testing.

Keywords: decision making; utilities; pseudodiagnostic reasoning

Introduction

A sales manager advertised a new position for a sales assistant. After reviewing the curriculum vitae of the candidates, she selects two promising applicants, Ms. A. and Ms. B. The manager initially leans towards the first candidate, Ms. A., and discovers that she successfully completed 70% of her sales transactions in the last month in her previous position. The manager must now engage in inductive reasoning: She needs to collect more information in order to decide whether Ms. A or Ms. B. is the best candidate for the job. A long tradition of psychological research suggests that her search for information will be driven by a need for evidence confirming the hypothesis she is entertaining. Thus, if, at this point, the manager believes Ms. A. *is* the best candidate, she would naturally seek more information about Ms. A rather than checking Ms. B.'s sale performance. Yet, this strategy is potentially shortsighted: Ms. B. could well have outperformed Ms. A. on the sales front in her previous job, in which case, seeking more information about Ms. A. could be misguided and lead to the employment of a candidate with less potential. Without establishing the sales performance of Ms. B, the diagnostic value of Ms. A's sales performance is undetermined, and hence cannot judiciously inform the decision-making process.

More generally, the diagnosticity of a datum D for a given hypothesis X (H_X) is defined in terms of the ratio of the probability that D is observed given that H_X is true, $P(D|H_X)$, and the probability that D is observed given that an alternative hypothesis Y is true, $P(D|H_Y)$. Hence, diagnosticity can only be assessed from the perspective of *multiple* hypotheses. The likelihood ratio in Bayes's Theorem is the normative metric of the diagnosticity of information (Doherty, Mynatt, Tweney, & Schiavo, 1979).

People's Search for Information Is Not (Always) Driven by Considerations of Diagnosticity

Early research examining how people gather information in order to make inferences suggested they did not fully appreciate that diagnosticity is defined in terms relative to at least two hypotheses, not just one (Beyth-Marom & Fischhoff, 1983; Doherty et al. 1979; Kern & Doherty, 1981; cf. Trope & Bassok, 1982, 1983). For example, Beyth-Marom and Fischhoff (1983, Experiment 1) told participants that an individual possessed a distinguishing feature and asked them what information they deemed relevant to determine whether that individual was a member of Group A. Nearly 90% of their participants indicated that it was relevant to know P(D|group A), but of those 'only' 50% deemed important to know the probability that this information would also be observed given membership in a different group, or P(D|group B). Yet both probabilities must be examined in order to gauge the diagnosticity of the distinguishing feature.

Mynatt, Doherty and Dragan (1993) argued that people's hypothesis testing process is predominantly concerned with gathering evidence in favour of one hypothesis, rather than determining the diagnosticity of any given piece of information for multiple hypotheses. In one of their reasoning scenarios, participants were asked to determine which of two cars, X or Y, their "sister" purchased. They were told about *two features characterising this car*, its petrol consumption (D_1: "25 miles per gallon" –mpg) and its mechanical reliability (D_2: "no major mechanical problems in the first two years of ownership"). In addition, participants were given *an anchoring piece of information*: "65% of car Xs do over 25 mpg", or $P(D_1|H_X) = .65$. Participants were then asked to choose which of the following three pieces of information would help them

decide which type of car was owned by their sister (the participants did not see the information presented here in brackets).

1. The percentage of car Ys that do over 25 mpg. [Diagnostic, $P(D_1|H_Y)$].
2. The percentage of car Xs that have had no major mechanical problems for the first two years of ownership. [Pseudodiagnostic, $P(D_2|H_X)$].
3. The percentage of car Ys that have had no major mechanical problems for the first two years of ownership. [Switching, $P(D_2|H_Y)$].

The first choice would establish the diagnosticity of the petrol consumption data. If more than 65% of car Ys do over 25 mpg, then the sister's car is more likely to be a Y car. Otherwise, it is more likely to be an X car. The second choice would determine the mechanical reliability of car Xs. This choice leaves the reasoner with two pieces of information, and the diagnosticity of neither can be established. Learning that a high percentage of car Xs also featured good mechanical reliability could make one confident that the sister owned a car X, but this feeling of confidence would only be illusory: here again, car Xs could well be outperformed by car Ys. Hence, choosing to learn about $P(D_2|H_X)$ is considered a pseudodiagnostic choice. We term the third choice *switching* because the focus switches from the initial information (D_1) to the second piece of information (D_2) and from the initial hypothesis (H_X) to the alternative one (H_y).

In Experiment 1 of Mynatt et al. (1993), 60% of the participants chose to learn about the percentage of car Xs with good mechanical reliability in order to determine the identity of the car, while only 26% chose to know the percentage of car Ys that do 25 mpg. The majority of participants thus made what is considered a pseudodiagnostic choice, since opting to look at the mechanical reliability of car Xs cannot determine the diagnosticity of being informed that 65% of car Xs do over 25 mpg.

In contrast, there is evidence demonstrating that people *will*, under various circumstances, seek to know information that would establish the diagnosticity of the anchoring information. For example, if the anchoring information defines a relatively rare feature, for example 65% of cars of make X have a top speed of 165 miles per hour, participants are more likely to want to know the proportion of the alternative make of cars that reach that top speed (Feeney, Evans, & Venn, 2008, Experiment 1). The rare, and arguably more interesting feature (it's plausibly more interesting to be told about a top speed of 165 mph than that the car has an ashtray), thus invite participants to gauge its frequency given the alternative category, thereby encouraging diagnostic data selection. In addition, if the dimensions that define the target object (e.g., 'your sister's car') are couched in terms of an actor's motivation (e.g., you sister *wanted* a car with good petrol consumption) then anchoring information that appears to run counter to this motivation (e.g., car X does 20 miles per gallon) elicits *less* pseudodiagnostic reasoning. Likewise, a low value for the initial anchoring information (e.g., 35% of car Xs do over 25 mpg) encourages more diagnostic choices (Mynatt et al., 1993, Experiment 2). Thus, upon learning that $P(D_1|H_X)$ is relatively low, more participants are interested in $P(D_1|H_Y)$. Conversely, if the anchoring information given plausibly endorses the focal hypothesis, that is when $P(D_1|H_X)$ is high, participants appear less motivated to determine the diagnosticity of that information.

Information Relevance and Initial Values in Action Problems

Another important characteristic that seems to determine whether people will make diagnostic search choices is the goal of the task. Mynatt et al. (1993) distinguished between *inference* and *action* problems. The car example discussed above, they argued, represents an inference problem. The car has already been purchased and is owned by someone, the goal is to determine whether it is a car X or a car Y. In effect the problem is a categorization inference, and in principle the categorization can be true or false. In contrast, an action problem is one where hypotheses represent two courses of action. One might be better than the other, but the decision cannot in principle be evaluated in terms of whether one action is true and the other false. In a separate experimental condition, Mynatt et al. instructed participants to imagine *buying* a car, considering car X or car Y and told them they were "concerned about (…) petrol consumption and mechanical reliability" (p. 768). Participants were then given the same anchoring piece of information ("65% of car Xs do over 25 mpg") and were asked to choose one among the same three pieces of information in order to help them decide which car to buy (see options 1. through 3. above).

In that action problem, 52% chose the piece of information that could determine the petrol consumption of car Ys (the diagnostic choice) and 41% chose the piece of information that could determine the mechanical reliability of car Xs (a pseudodiagnostic choice). To explain the high proportion of diagnostic choices in action problems, Mynatt et al. propose that the choice among the three alternatives is determined by the datum which bears more utility for each individual participant: "Precisely how many subjects will select (the diagnostic choice) will depend on the content of a given problem and subjects' *idiosyncratic* utility function and decision strategies" (pp.765-766, emphasis added). On this account, those who consider petrol consumption to be more important than mechanical reliability would be motivated to establish the petrol consumption of car Ys and hence chose the diagnostic option. In contrast, those who are more concerned about mechanical reliability should seek information about car Xs' mechanical reliability, a pseudodiagnostic choice.

The authors, however, did not manipulate explicitly the perceived relevance of the two dimensions characterising each alternative (e.g., petrol consumption and mechanical reliability in the car scenario) nor did they seek to assess how relevant their participants believed these dimensions to be. Moreover, there is conflicting evidence showing that action problems may not necessarily promote diagnosticity. Maggi, Butera, Legrenzi and Mugny (1998, Experiment 1) asked participants to imagine having to choose between two cars or two political candidates. These authors found over 60% choices to be pseudodiagnostic. There was, however, an important methodological difference between their task and that of Mynatt et al. (1992): Maggi et al. (1998) presented participants with *four* possible pieces of information to choose from for each alternative (e.g., the car price, reliability, fuel consumption and performance). In addition, the authors found that people tended to be more diagnostic in their choices when the anchoring information concerned a characteristic they believe to be important (e.g., the price of a car or the competence of a political candidate). In light of these incongruous findings, one important issue to resolve would thus be to determine whether those who made a diagnostic choice by choosing to look up $P(D_1|H_Y)$ did so *because* they were more interested in D_1 than in any other piece of information D.

Another important difference between inference and action problems outlined by Mynatt et al. (1993) is the role of the initial $P(D_1|H_X)$ value of the anchoring information. According to the authors, in inference problems this initial value could be a cue to the truth value of H_X and, as such, dictate participants' information search. By contrast, the authors predicted and found that this initial value would not affect choices in action problems since, in those situations, information search would solely be determined by the perceived relevance of the anchoring information D_1. The authors, however, tested this prediction by comparing relatively narrow values, hovering modestly below and above the 50% mark (viz. 35% vs. 65%). It is therefore reasonable to assume that participants who believed, for example, petrol consumption was *the* most important attribute for a new car would always wonder if car Ys outperformed car Xs, even upon learning that 65% car Xs did over 25 mpg. However, this does not necessarily imply that participants' information search strategy will never be affected by the $P(D_1|H_X)$ value in action problems. It is not implausible to expect, for example, that participants may no longer search to establish a diagnosticity ratio when told that 95% of car Xs do over 25 mpg. In this case, the $P(D_1|H_X)$ value could be a cue to the superiority of H_X. Consequently, under such circumstances, participants might then be more interested in learning more about car Xs than in finding out whether car Ys outperform car Xs. Hence, when the $P(D_1|H_X)$ value is deemed satisfactory in action problems, we should expect more pseudodiagnostic choices.

The Present Study

In the experiment reported here, we examined the role of information relevance and initial values in determining diagnostic and pseudodiagnostic choices in problems structurally isomorphic to the one developed by Mynatt et al. (1993). The first aim of this experiment was to test the hypothesis that diagnostic choices in action problems occur because people believe D_1 is more relevant than D_2 in deciding whether to take action X or action Y. We manipulated the relative importance of D_1 and D_2 in two scenarios so that participants would care more about D_1 than D_2. We anticipated the higher perceived relevance of D_1 would lead to a higher proportion of diagnostic choices, since participants would seek to determine the probability of D_1 given the alternative course of action. Second, this experiment aimed to assess the degree to which people may revert to a pseudodiagnostic search for information when the initial value of the anchoring information $P(D_1|H_X)$ is deemed satisfactory. To do so, we manipulated the motivation underpinning participants' action. One scenario was designed to motivate people to find *the highest value* of $P(D_1|H)$. In this case, we anticipated that participants would never be satisfied by the initial value of $P(D_1|H_X)$ and hence we predicted that their search for information would not be affected by this initial value. The alternative scenario was designed to motivate people to find *a satisfactory value* of $P(D_1|H)$. In such a situation, we predicted that when the initial $P(D_1|H_X)$ presented could be deemed satisfactory, the rate of pseudodiagnostic choices would be greater.

Method

Participants

Participants were recruited by third-year psychology students at the University of Toulouse, France, as a course requirement. Each student made a list of several men and women who were older than 18 and not studying psychology, randomly drew one man and one woman from his or her list, and asked them to take part in a general survey which included the present study. Of the 1040 participants in the final sample (520 men, 520 women; mean age = 31.37, *SD* = 13.24), 11% had completed graduate school, 53% had an undergraduate education, 20% had graduated from high school only, and the remaining 16% had not graduated from high school. The sample included a large proportion of students (40%), but also working professionals (51%) and retired or unemployed individuals (8%). The survey was conducted in French.

Design and Procedure

The current experimental manipulation was embedded in a longer questionnaire. The experiment used a 2 × 2

between-subjects design. The independent variables were the implicit motivation of the decision-maker (maximizing vs. satisficing – Simon, 1955) and the numerical value of the anchoring piece of information (high or low). Participants were randomly allocated to one of the resulting four conditions. Their task was presented as follows: they were asked to imagine they were the director of a large zoo and that they had set up a programme aiming to promote reproduction in captivity of African elephants, a species at risk of extinction. Their calves, however, were facing a severe health issue. They were informed of the presence of a parasite whose eggs could lodge in the calves' aortic artery, causing strokes and killing the calves if left untreated, threatening the success of the reproduction programme. The experimental manipulation concerned the animals needed to be treated in order to rid the zoo of parasites. Half the participants were told the deadly parasites were infecting the calves directly and treatment was, therefore, to be administered to calves (satisficing scenario). The remaining half was told the deadly parasites were carried by roaming rats which were to be treated directly (maximizing scenario). In all cases, the zoo's chief veterinary suggested using one of two treatments to save the calves: treatment A or treatment B. Participants were then told about the mortality rate of calves (rats) treated with treatment A and that both treatments could also potentially cause infertility in calves (rats). The initial value of the anchoring piece of information was either high or low. Thus, half were told treatment A could cause the death of 80% calves (rats) whereas the remaining half were told it could cause the death of 20% calves (rats). Before making their choice, however, they were allowed to consult one additional piece of information among three alternatives: they could choose to consult the mortality rate of calves (rats) treated with treatment B (a diagnostic choice). They could also choose to learn more about treatment A and ask to consult the percentage of infertile calves (rats) among those treated with treatment A (a pseudodiagnostic choice). Finally, they could choose to learn about the rate of infertility observed in calves (rats) treated with treatment B (a switching choice). The order of the choices remained constant in all experimental conditions.

In both scenarios, we anticipated that people would be more concerned about mortality rates (D_1) than about infertility rates (D_2). Specifically, we anticipated people to place more value on the mortality rate of rats than on their infertility since rats made infertile would not eliminate their status as a contamination vector. Likewise, we anticipated people would be more concerned about avoiding the death of the endangered calves than about their potential infertility. A few pages later in the survey, all participants were asked to consider the calves (rats) task again and to rate the importance of avoiding killing the calves (rats) as well as the importance of avoiding making the calves (rats) sterile. Both ratings were recorded on an

8-point scale ranging from 1 (Absolutely not important) to 8 (Extremely important).

We expected that the greater relevance of the anchoring dimension (the mortality rate) induced by the scenarios would result in a large proportion of diagnostic choices. In the maximizing scenario, we predicted that participants would be motivated to find the best treatment to kill *all* the rats, and that consequently, short of a 100% mortality rate, they would be more interested in determining the mortality rate of the alternative treatment *regardless* of the mortality rate for treatment A. As a result, we expected high proportions of diagnostic choices in these scenarios when the mortality rate for treatment A was set at either .20 or .80. In the satisficing scenario, we anticipated that mortality would bear unacceptable consequences to a degree that varied with the rate associated with treatment A. We predicted that participants would deem it important to save as many calves as possible and that, consequently, the 80% chance of killing the host organism associated with treatment A would be deemed unacceptably high. In this situation, we predicted a strong preference for enquiring about the mortality rate associated with treatment B, the diagnostic option. With a lower mortality rate of .20, however, we predicted that some participants might deem it satisficingly low and hence be tempted to enquire about the infertility rate associated with treatment A, resulting in higher proportion of pseudodiagnostic choices in this condition.

Table 1: Mean and standard deviation of the importance ratings of preventing killing or preventing infertility in the host organism (rats or calves) in the conditions where the rats and elephant calves are treated as a function of the mortality rate for treatment A, .2 and .8.

		Motivation			
		Maximizing (rats)		Satisficing (calves)	
		Mortality Rate of Treatment A		Mortality Rate of Treatment A	
		.2	.8	.2	.8
Prevent Killing	M	3.37	3.59	7.19	7.18
	SD	2.30	2.48	1.17	1.32
Prevent Infertility	M	3.09	3.30	6.93	6.87
	SD	2.32	2.57	1.36	1.64

Results

Importance Ratings

Participants were asked to rate the importance of not killing the host animals or not making them infertile. These mean importance ratings are reported in Table 1. The importance of saving the calves was consistently rated higher than that of saving the rats. The same was true for the infertility ratings: Participants judged it more important to prevent infertility in the calves than in the rats. Most notably, as we had anticipated, participants deemed it more important to prevent mortality than infertility.

A 2 (goals: maximizing, satisficing) × 2 (mortality rate of treatment A: .2, .8) × 2 (rating type: killing, infertility) mixed analysis of variance (ANOVA) confirmed these observations. The main effect of goal was significant, $F(1, 1036) = 1255$, $p < .001$, $MSE = 5.69$, $\eta^2 = .55$, the main effect of the mortality rate was not significant, $F < 1$, and the main effect of rating type was significant, $F(1, 1036) = 20.1$, $p < .001$, $MSE = 2.11$, $\eta^2 = .02$. None of the interactions were significant, largest non reliable $F(1, 1036) = 1.46$.

Choice Preferences

Two participants did not make a choice and were discarded from subsequent analyses. Consistent with our predictions, the diagnostic option was by far the most frequently chosen in all experimental conditions with over 70% of participants opting for this type of information (see Fig. 1). In the maximizing scenarios (treating the rats), nearly 80% of the participants elected to examine the mortality rate associated with treatment B, and the remaining 20% of the participants were evenly split between the other two options (the infertility rate for treatment A or B). Moreover, this pattern of choice was identical whether the value of the mortality rate of treatment A was said to be 20% or 80%. In contrast, in the satisficing scenarios (treating the calves), while most participants still elected primarily to determine the mortality rate of treatment B (the diagnostic choice), the frequency of pseudodiagnostic choices was nearly twice as large when treatment A had a relatively low mortality rate (20%) compared to when it had a high mortality rate (80%). Approximately one fifth of the participants chose the irrelevant option, the infertility rate of treatment B in both versions of the satisficing scenario.

A number of χ^2 tests were conducted. The first determined that the choice frequencies in all four experimental conditions differed significantly, χ^2 (df = 6, $N = 1038$) = 35.2, $p < .001$. The proportion of diagnostic choices was significantly higher when the implicit goal was to maximize the number of rats killed, χ^2 (df = 1, $N = 869$) = 5.46, $p < .02$. Separate tests were then conducted within the maximizing (rats) and the satisficing (calves) scenarios, excluding the switching choice frequencies. Within the maximizing scenarios, the frequencies of diagnostic and pseudodiagnostic choices did not differ as a function of the mortality rate of treatment A, χ^2 (df = 1, $N = 463$) = .11, $p > .05$. In contrast, within the satisficing scenarios, the frequencies differed significantly as a function of the mortality rate of treatment A, χ^2 (df = 1, $N = 406$) = 9.22, $p < .005$.

Discussion

This experiment successfully manipulated the relative importance of the information dimensions available in a two-alternative action problem where participants were asked to choose which treatment they should use to save endangered calves whose life was threatened by a deadly parasite. Specifically, ratings of the importance of not killing the animals and not making them infertile confirmed that the mortality dimension was judged more important than the infertility dimension whether the animals were the calves themselves or rats hosting the parasite.

The examination of information search patterns in turn confirmed that when the anchoring dimension was perceived as being the most relevant, participants were strongly drawn to check the diagnostic option: nearly 71% of the 1038 responses collected were diagnostic. This high
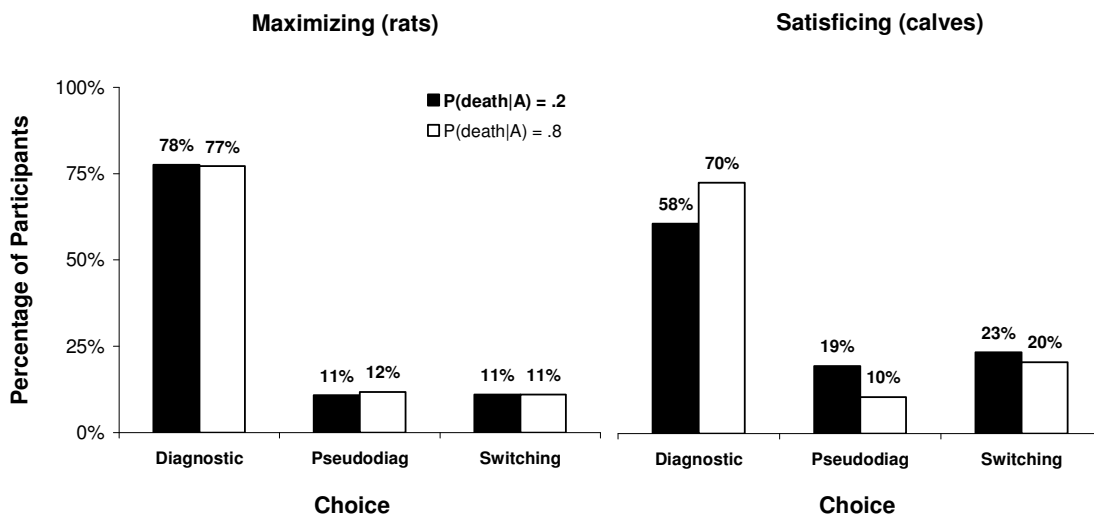

Figure 1: Percentage of participants making a diagnostic, pseudodiagnostic, or switching choice for the maximizing scenarios (involving the rats) and the satisficing scenarios (involveing the calves) for both values of P(death|treatment A).

degree of consensus is all the more impressive given the size and the variety of the sample from which it originated. This result supports Mynatt et al.'s (1993) initial, albeit untested, assumption that individuals' search for information is driven by considerations of utility in action problems.

We note the large inconsistency between the rate of diagnostic choice observed in our action problems (on average, more than 70% of our participants chose the diagnostic option) compared to those observed by Maggi et al. (1998) with a similar task (on average, less than 40% of their participants did so). These authors, however, had also found that people were more diagnostic when the anchoring dimension was one they judge to be important. Recall that an important methodological difference between their task and the original action problems used by Mynatt et al. (1993), as well as that used in the present study, was the number of dimensions participants could choose from. Whereas our participants and Mynatt et al.'s (1993) could only choose to look up the probability associated with two dimensions (D_1 and D_2) for each of two alternatives, Maggi et al.'s participants were presented with four such dimensions (D_1, D_2, D_3, and D_4). Moreover, the authors rotated the dimension defined as the anchoring dimension so that some participants would be first given information about D_1, others about D_2 and so on. Suppose that participants' search strategy is primarily driven by the importance of the dimension and suppose D_1 was the dimension they deemed most important. This means that whenever the anchoring dimension was *not* the most important dimension (3 times out of 4), participants would seek to learn more about D_1 for the current alternative and hence make a pseudodiagnostic choice. In other words, perhaps the reason why so many people made pseudodiagnostic choices in Maggi et al.'s (1998) task was because most of the time the anchoring dimension was not the dimension bearing the highest utility.

Finally, in line with our initial predictions, but contrary to Mynatt et al.'s (1993) conclusions, we were able to demonstrate that the numerical value of the anchoring dimension could affect people's search strategies when their motivation was to find a satisficing alternative. In such circumstances, a more satisficing value (viz., a relatively low mortality rate of endangered calves) resulted in almost twice as many pseudodiagnostic choices than a plainly unsatisfactory value (viz., a high mortality rate). This suggests that people will *also* engage in confirmatory search for information when they aim to choose between two courses of actions (and not only when they seek to make an inference, as the authors had previously concluded). These data therefore offer strong support for the hypothesis that the perceived relevance of the dimensions that define two courses of action governs the information search strategies adopted by reasoners. In addition, they establish that such strategies can also be modified depending on what the decision-maker is motivated to achieve, namely either identify a satisficing alternative or identify a utility maximizing alternative.

References

Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudo-diagnosticity. *Journal of Personality and Social Psychology*, *45*, 1185-1195.

Doherty, M. E., Chadwick, R., Garavan, H., Barr, D., & Mynatt, C. R. (1996). On people's understanding of the diagnostic implications of probabilistic data. *Memory and Cognition*, *24*, 644-654.

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica*, *43*, 111-121.

Feeney, A., Evans, J., Venn, S. (2008). Rarity, pseudodiagnosticity and Bayesian reasoning. *Thinking & Reasoning, 14*, 209-230.

Kern, L., & Doherty, M. E. (1982). "Pseudodiagnosticity" in an idealized medical problem-solving environment. *Journal of Medical Education*, *57*, 100-104.

Maggi, J., Butera, F., Legrenzi, P., & Mugny, G. (1998). Relevance of information and social influence in the pseudodiagnosticity bias. *Swiss Journal of Psychology, 57*, 188-199.

Mynatt, C. R., Doherty, M. E., & Dragan, W. (1993). Information relevance, working memory, and the consideration of alternatives. *Quarterly Journal of Experimental Psychology*, *46A,* 759-778.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99-118.

Trope, Y., Bassok, M. (1982). Confirmatory and diagnosing strategies in social information gathering. *Journal of Personality and Social Psychology*, *43*, 22-34.

Trope, Y., Bassok, M. (1983). Information-gathering strategies in hypothesis-testing. *Journal of Experimental Social Psychology, 19*, 560-576.