

THEORETICAL NOTE

Pseudodiagnosticity Revisited

Vincenzo Crupi

University IUAV of Venice and University of Turin

Katya Tentori and Luigi Lombardi

University of Trento

In the psychology of reasoning and judgment, the pseudodiagnosticity task has been a major tool for the empirical investigation of people's ability to search for diagnostic information. A novel normative analysis of this experimental paradigm is presented, by which the participants' prevailing responses turn out not to support the generally accepted existence of a reasoning bias. The conclusions drawn do not rest on pragmatic concerns suggesting alleged divergences between the experimenter's and participants' reading of the task. They only rely, instead, on the demonstration that observed behavior largely conforms to optimal utility maximizing information search strategies for standard variants of the pseudodiagnosticity paradigm that have been investigated so far. It is argued that the experimental results obtained, contrary to what has recurrently been claimed, have failed to discriminate between normative and nonnormative accounts of behavior. More general implications of the analysis presented for past and future research on human information search behavior and diagnostic reasoning are discussed.

Keywords: pseudodiagnosticity, Bayesian reasoning, epistemic utility, rationality

Active search for information from the environment is an essential part of cognition. In human affairs, looking for additional information is often of crucial importance, as decision problems with limited evidence initially available are commonplace. Indeed, when competing hypotheses have to be compared, relevant information often has to be identified and selected from a huge amount of potentially available data. As resources (e.g., time, money) are typically scarce, moreover, information search has to be selective. It is therefore important to understand whether people's information search strategies allow them to identify highly diagnostic items out of all the accessible data.

As a matter of fact, the search for evidence in order to evaluate hypotheses has been seen as a central issue for the study of thinking, with important implications in applied settings (Baron, 2000, pp. 4–15). Medical practice is a case in point. When facing a clinical problem, the need often arises to acquire additional information under resource constraints (e.g., performing some diagnostic tests but forgoing others). The effectiveness of treatment decisions, then, clearly rests on the physician's ability to selectively search for relevant clinical evidence on the basis of an appropriate assessment of how it relates to the diagnostic hypotheses under consideration.

In the epistemological, psychological, and medical literature, Bayes's theorem is often assumed as a convenient framework to assess the diagnosticity of an observed datum or piece of evidence relative to a hypothesis set (see Fischhoff & Beyth-Marom, 1983; Good, 1950; Sackett, Haynes, & Tugwell, 1985). In its well-known odds form, the theorem can be written as follows (P stands for a probability function and “ \neg ” for logical negation):

$$\frac{P(h|e)}{P(\neg h|e)} = \frac{P(h)}{P(\neg h)} \times \frac{P(e|h)}{P(e|\neg h)} \quad (1)$$

Reading from the left, the three terms in the equation are as follows: the *posterior odds* of hypothesis h (vs. $\neg h$) being true given all that is known after the receipt of information e ; the *prior odds* of h (vs. $\neg h$) being true given all that is known before the receipt of e ; and the *likelihood ratio* of h (vs. $\neg h$) relative to e . The likelihood ratio is commonly used to capture the diagnostic value of evidence e relative to h versus $\neg h$. More precisely, (a) if $P(e|h)/P(e|\neg h) > 1$, then e provides support for h against $\neg h$, for h surpasses $\neg h$ as a predictor of e ; as a consequence, the posterior odds of h ($\neg h$) will be higher (lower) than the prior ones; (b) if $P(e|h)/P(e|\neg h) < 1$, then e provides support for $\neg h$ against h , for $\neg h$ surpasses h as a predictor of e ; as a consequence, the posterior odds of h ($\neg h$) will be lower (higher) than the prior ones; (c) if

Vincenzo Crupi, Department of Arts and Design, University IUAV of Venice, Venice, Italy, and Department of Philosophy, University of Turin, Turin, Italy; Katya Tentori, Center for Mind/Brain Sciences (CIMeC) and Department of Cognitive Sciences and Education, University of Trento, Trento, Italy; Luigi Lombardi, Department of Cognitive Sciences and Education, University of Trento.

This research was supported by a grant from the SMC/Fondazione Cassa di Risparmio di Trento e Rovereto for the CIMeC (University of Trento) research project on *Inductive reasoning*. Tommaso Mastropasqua provided invaluable help with the mathematical treatment and a number of insightful suggestions on the text. Comments and contributions from Paolo Cherubini, Michel Gonzalez, and Massimo Vescovi have also been critical to the completion of this work. Nick Chater, Anna Maria Cherubini, Jonathan Evans, Branden Fitelson, Vittorio Girotto, Paolo Legrenzi, Dan Osherson, and Mike Titelbaum must be credited for useful remarks and exchanges. Enrico Michelotti helped us with computer programming during early stages of this research work, and Valentina Montresor composed the figures appearing in the text. Finally, we would like to express our gratitude to Michael Doherty for his kindness and willingness to discuss the content and consequences of this work.

Correspondence concerning this article should be addressed to Vincenzo Crupi, Department of Philosophy, University of Turin, via Sant'Ottagio 20, 10124 Torino, Italy. E-mail: crupi@iuav.it or vincenzo.crupi@unito.it

$P(e|h)/P(e|\neg h) = 1$, then e is equally consistent with both h and $\neg h$, as neither of them predicts e better than the other; as a consequence, the posterior odds remain unchanged relative to the prior ones. In Cases (a) and (b), i.e., when the likelihood ratio departs from 1) evidence e is said to carry some diagnostic value, whereas in Case (c), e makes no contribution in discriminating h and $\neg h$, thus having no diagnostic value relative to the hypothesis set at issue.

More or less elaborated measures have been devised to quantify to what degree the likelihood ratio departs from 1, thus how large is the impact and diagnostic value of a given e in comparing the credibility of hypothesis h versus $\neg h$ (see Crupi, Tentori, & Gonzalez, 2007; Fitelson, 2001; Good, 1950). Notably, a highly diagnostic piece of evidence e having been acquired does not necessarily imply that the probability of one hypothesis becomes high and that of the alternative low, for posterior odds also depend on the prior ones. High diagnosticity means only that the probability of h versus $\neg h$ significantly changes as a consequence of coming to know e . Also notice that a piece of evidence e can strongly support hypothesis h even if $P(e|h)$ is rather low, provided that e is even more unlikely given $\neg h$. Symmetrically, e can significantly support $\neg h$ even if $P(e|h)$ is quite high, provided that e is even more likely given $\neg h$.

People's ability to appropriately appreciate diagnosticity has been claimed to be severely limited on the basis of psychological findings involving a rather widespread experimental procedure, i.e., the "pseudodiagnosticity" paradigm. According to Doherty, Chadwick, Garavan, Barr, and Mynatt's (1996) survey, this paradigm amounts to one of the two most popular experimental devices explicitly devoted to the investigation of people's understanding of the diagnostic implications of data (the other being represented by the "Vuma planet" scenario, originally introduced by Skov & Sherman, 1986, to be discussed later on along with other studies outside the pseudodiagnosticity paradigm). It is also the one that has attracted relatively more attention and effort, presumably because human performance on the pseudodiagnosticity task has been seen as distinctively poor.

Our present aim is to question the commonly accepted analysis of the pseudodiagnosticity task as normatively unsound and to replace it with a more appropriate treatment. Notably, we do not mean to contend the Bayesian framework; rather, we will argue that it has been importantly misapplied. Our contribution is theoretical in nature, with a range of implications for empirical research. First, as the psychology of reasoning still exhibits lively debated concerns about diagnoses of human rationality versus irrationality, it remains of interest to scrutinize the tenability of such diagnoses. It will be argued that, on the basis of a misguided reading of the task, pseudodiagnosticity results prompted hasty conclusions on people's allegedly poor performance in information search, thus misleadingly contributing to belief in so-called confirmation bias and the perception of its widespread consequences. Second, it will be shown that a number of relevant studies exhibit shortcomings in experimental design, methods, or results reporting that emerge from a lack of theoretical analysis. Future research on humans' understanding of diagnosticity is thus expected to significantly profit from having a rigorous treatment of the pseudodiagnosticity task at hand.

The rest of the article is organized as follows: After reviewing the experimental evidence available and its usual interpretation, we

will argue that the task of the pseudodiagnosticity paradigm has not been analyzed thoroughly enough. Upon closer scrutiny, observed behavior in standard versions of this task turns out to be largely consistent with a novel compelling normative analysis of the problem. Inquiries with other variants of the task occurring in the pseudodiagnosticity literature will also be discussed and found to have reported results that are inconclusive in light of the current analysis. Overall, thus, the results available to date do not convincingly support the pessimistic conclusion that people fall short of a sound assessment of diagnosticity. As pointed out in the subsequent discussion, however, this does not imply the opposite conclusion either, i.e., that human agents generally evaluate diagnostic value in an optimal fashion. We will argue instead that future tests of people's ability to assess diagnosticity will need to be conceived as clearly discriminating between appropriate models of rational behavior and nonnormative accounts of human reasoning.

Reviewing Pseudodiagnosticity

The Experimental Paradigm

In the pseudodiagnosticity task, participants are presented with two mutually exclusive and jointly exhaustive hypotheses, h versus $\neg h$ (e.g., two possible illnesses, one and only one of which is said to affect a certain patient). The comparison between the hypotheses involves two different data f and g already observed (e.g., the presence of two symptoms in the patient considered). The probabilistic relationships between the available data (f and g) and the hypotheses under consideration (h vs. $\neg h$) can then be represented in a 2×2 array as shown in Table 1. Participants are informed that precise values are available for the four likelihoods $P(f|h)$, $P(f|\neg h)$, $P(g|h)$, and $P(g|\neg h)$, reflecting expected frequencies in relevant samples. $P(f|h)$, for instance, would amount to the expected rate of occurrence of symptom f in a suitable sample of patients for which diagnosis h holds. In fact, participants are initially provided with one such likelihood; for instance, they are given the value of $P(f|h)$, as seen in Table 1. Then they are told that they can ask for only one among the three still-concealed values in the table. Their task is to indicate which value they would prefer to be given in view of a decision about h versus $\neg h$. Participants are thus instructed to select what they see as the additional piece of information most useful to the ultimate goal of a subsequent choice between competing hypotheses.

In an effort to control for possible confounds, most researchers have chosen hypotheses h and $\neg h$ and data f and g so as to discourage participants from elaborating any specific conjecture concerning the undisclosed likelihood values. For instance, should f and g bear an appreciably strong causal connection, then a

Table 1
A Schematic Illustration of Pseudodiagnosticity Scenarios

Data	Hypotheses	
	h	$\neg h$
f	Likelihood provided	Likelihood not provided
g	Likelihood not provided	Likelihood not provided

relatively high value of the initially given likelihood concerning f would suggest a relatively high value of the corresponding likelihood concerning g . Such an inference has typically been prevented by choosing f and g as indicating conditions which by and large appear to be independent (e.g., symptoms such as cough and leg pain). Moreover, to prevent participants from elaborating their own estimates of (ranges of most plausible) values for the unknown likelihoods on the basis of their background knowledge, researchers have usually presented hypotheses h and $\neg h$ as referring to abstract categories (e.g., disease A and disease B). As usual in the inquiry on human thinking, these cautions have been taken to let “the purest kind of reasoning” occur in the experimental setting (Kern & Doherty, 1982, p. 104).

We will label *standard pseudodiagnosticity task* the class of experimental problems in which the foregoing conditions are fulfilled. The standard version of the task includes a majority of instances in which h and $\neg h$ have been presented as having an equal initial probability (or base rate) of .50 and the anchor value provided has been ranging from moderately high (such as .58, in Wolf, Gruppen, & Billi, 1985, Case 2) to high (such as .84, in Kern & Doherty, 1982). Occasionally, low anchor values have been employed, with either equal or unequal base rates (see Mynatt, Doherty, & Dragan, 1993, Exp. 2, and Wolf et al., 1985, Case 3, respectively). These variants still belong to the standard version of the task as presently defined. It should be noted, however, that nonstandard versions have also been conceived, to deliberately let participants use their background knowledge to elaborate estimates for at least some of the initially unknown likelihoods (see Feeney, Evans, & Venn, 2000, 2008).

In our treatment, we will analyze standard versions of the pseudodiagnosticity task first. Nonstandard variants will be discussed afterward. Before all that, however, we will have to further survey the pseudodiagnosticity literature in terms of the results typically obtained, their commonly accepted interpretation, and the consequences drawn.

Usual Interpretation, Results, and Consequences

The usual interpretation of the pseudodiagnosticity task has been said to rest on Bayesian rationality, and specifically on the Bayesian notion of diagnosticity as described in the introduction above. In essence, the commonly accepted analysis runs as follows: First, it is claimed that appropriately revised probabilities of h and $\neg h$ “can be calculated only if *both* the probability of a piece of data given the hypothesis under test, $P(f|h)$, and the probability of the same piece of data given the alternative hypothesis, $P(f|\neg h)$, are known” (Doherty, Mynatt, Tweney, & Schiavo, 1979, p. 112). That is, in a Bayesian approach, information can be diagnostic only if it gives the likelihood ratio of two hypotheses (see Evans & Over, 1996b, p. 65; also see Manktelow, 1999, pp. 134 ff). Recall that initially participants have one likelihood value at their disposal (see Table 1) and are constrained to select only one further likelihood. It is pointed out that the value aligned with the given likelihood down the column in the table does not allow for the completion of any likelihood ratio. The same holds for the value aligned with the given likelihood along the diagonal. It is then asserted that such data are completely worthless in a rational approach to the experimental task. Finally, since the value aligned with the given likelihood along the row *does* allow for the com-

pletion of one of the likelihood ratios (i.e., the ratio between the probabilities of one single already known fact, f , under the two competing hypotheses h and $\neg h$), it is concluded that Bayesian standards prescribe the selection of the row value as rationally mandatory.

Experimentally observed behavior has shown a marked departure from such a prescription. The robust finding across a number of standard problems involving a relatively high anchor value and equal priors has been a widespread tendency to select the column value, i.e., precisely one of the allegedly worthless pieces of information. Coupling this kind of result with the normative reading summarized above, Doherty et al. (1979) concluded: “The subjects actively chose irrelevant information and ignored relevant information which was equally easily available” (p. 119). Notably, a different pattern of responses has been documented by Mynatt et al. (1993, Exp. 2) with a low anchor (and equal priors): With an initially given likelihood of .35, the majority of their participants did choose to search for the row value.

The tendency to select a cell value other than in a row along a relatively high anchor has been seen as a strong demonstration that in this information search task “the majority of subjects have little or no intuitive understanding of the implications of Bayes’s theorem” (Mynatt et al., 1993, p. 762). Such a behavior has thus been labeled *pseudodiagnosticity* and connected to the overarching pattern of so-called confirmation bias (Doherty et al., 1979; Klayman, 1995; Nickerson, 1998). Accordingly, efforts have been devoted to the experimental investigation of factors and interventions that may bring participants’ responses in better alignment with the one usually identified as rationally correct (see Evans, Venn, & Feeney, 2002, Exp. 2; Wolf, Gruppen, & Billi, 1988).

A major candidate explanation of the experimental findings obtained has been proposed in terms of biasing processes of focalization. According to this proposal, a high anchor value connecting f and h would typically establish h as the focal hypothesis to the detriment of the alternative $\neg h$, thus inducing many participants to find out something more about the former (the column value) rather than about the latter (the row value). On such a basis, Mynatt et al. (1993) even speculated that “people can only think of one hypothesis at a time” (p. 774). They postulated, moreover, that a low anchor value would typically shift the focus on $\neg h$, thus inhibiting the otherwise common selection of the column value, which would account for the observation of more choices for the optimal option (the row value) with a low anchor (see Evans, 2007, for a more recent and detailed discussion along similar lines).

Results of pseudodiagnosticity experiments have been seen as having rather far-reaching consequences in discussions on human performance, human rationality, and their limitations. It has been suggested that pseudodiagnostic behavior may contribute to erroneous inferences, thus nonoptimal actions, for instance in clinical practice (Doherty et al., 1996, p. 653). Indeed, pseudodiagnosticity has been claimed to represent a major class of errors in decision making by inexperienced physicians (Wood, 1999, p. 602). More generally, Fischhoff and Beyth-Marom (1983) have identified the tendency to ignore the likelihood of $\neg h$ when evaluating evidence for h as a “powerful metabias” in human reasoning (p. 257). Stich (1990) resorted to the pseudodiagnosticity phenomenon as a crucial example countering philosophical and evolutionary claims that irrationality is impossible (pp. 7 ff), and Dawes (2001) presented

findings on pseudodiagnostic behavior as fostering the conclusion that in fact irrationality is abundant in everyday life as well as in expert judgment (chaps. 1 and 5).

A New Analysis: Framework and Assumptions

Consider again the standard pseudodiagnosticity task as illustrated in Table 1. Is the selection of the row value really the optimal solution or even the only rational strategy (Doherty et al., 1979, p. 117)? These conclusions crucially rest on the assumption that this value is “the only information that will allow normatively appropriate computations” (Doherty et al., 1996, p. 645). Our major aim here is to question the latter claim by presenting a new formal analysis of the task. As a consequence, the above conclusions will be shown to remain unsupported.

This and the next section are arranged as follows: First, we will describe and analyze the goal that the experimental problem and instructions suggest. We will then identify and briefly expose the appropriate (Bayesian) normative model for information selection to fit the experimental problem and the relevant goal. After reviewing the probabilistic structure of the task, we will apply that normative model to it.

The treatment to follow will require a certain amount of formalism. For ease of recall, a comprehensive summary of the notation and symbols introduced and employed throughout the article appears in Table 2.

Expected Gain in Epistemic Utility

Consider a set of mutually exclusive and jointly exhaustive hypotheses. We will say that an agent acts as a truth seeker iff she simply attributes utility 1 to the choice of the true hypothesis and

utility 0 to the choice of any false hypothesis. A truth seeker’s utilities can be seen as epistemic in the sense that the agent cares only for capturing the true state of the world in a given domain of inquiry as defined by the partition of hypotheses at issue.

In the pseudodiagnosticity paradigm, the partition simply amounts to complementary hypotheses h and $\neg h$. A truth-seeking agent would then assess utilities as illustrated in Table 3.

Recall that participants are asked to indicate their preferred information search in view of the ultimate aim of choosing h versus $\neg h$. Also, they are usually provided with no hint whatsoever as to how desirable hitting h would be as compared with hitting $\neg h$ or as to how serious missing h would be as compared with missing $\neg h$. To that extent, we claim that participants are encouraged to act as truth seekers. Notably, this assumption is in line with the remark explicitly made by Mynatt et al. (1993, p. 773) that the pseudodiagnosticity paradigm controls for motivational factors, as no differential payoff for confirming or disconfirming either h or $\neg h$ is present.

For a truth seeker, the expected utility of choosing any hypothesis h simply amounts to the probability of that hypothesis. In fact,

$$EU(\text{choosing } h) = [P(h) \times U(\text{hitting } h)] + [P(\neg h) \times U(\text{missing } \neg h)] = [P(h) \times 1] + [P(\neg h) \times 0] = P(h). \quad (2)$$

When choosing among hypotheses, a rational truth seeker will simply select the most probable one, thus maximizing expected utility. How should a truth seeker behave when searching for information? The short answer is: as any utility-maximizing agent would. When a choice is needed, she should always prefer and pursue the search for information providing the highest gain in expected utility. For a truth seeker, such a gain amounts to the difference between the expected probability of the hypothesis that will be chosen after the (still unknown) piece of information has been eventually acquired and the probability of the hypothesis that would be presently chosen (i.e., without any search for information being pursued).

The general formal model expressing this measure of the value of an information search has been presented by Baron (1985, chap. 4), following Savage (1954, chap. 6). Let H be a finite set of mutually exclusive and jointly exhaustive hypotheses at issue. Then let X be the set of possible alternative values of a variable of interest. Our present point is the evaluation of the expected utility gain of finding out the (still unknown) true value within X for a choice within hypothesis set H .

Elaborating on Equation 2, a truth seeker’s expected utility for choosing a hypothesis within set H without further inquiry, here denoted as $EU(H)$, equals simply the probability of the most likely hypothesis in the set, that is,

$$EU(H) = \max_{h_i \in H} (P(h_i)). \quad (3)$$

By a straightforward extension of Equation 3, the expected utility of choosing a hypothesis after finding out a given value $x \in X$, denoted as $EU(H|x)$, amounts to the probability of the most likely hypothesis within H conditional on knowing x , that is,

$$EU(H|x) = \max_{h_i \in H} (P(h_i|x)). \quad (4)$$

Finally, the expected utility of a choice subsequent to an information search about X , calculated when the sought piece of informa-

Table 2
Notation and Symbols Employed

Notation or symbol	Definition
$EU(H)$	Expected utility of a choice in hypothesis set H
$EU(H x)$	EU of a choice in H given evidence x
$EU_X(H)$	EU of a choice in H following a (yet to be performed) search about X
$EU_X(H y)$	EU of a choice in H following a (yet to be performed) search about X given evidence y
$\Delta EU(H, X)$	EU gain associated to a (yet to be performed) search about X for a choice in H
$\Delta EU(H, X y)$	EU gain associated to a (yet to be performed) search about X for a choice in H given evidence y
$h, \neg h$	Hypotheses at issue in the pseudodiagnosticity task
$H = \{h, \neg h\}$	Hypothesis set in the pseudodiagnosticity task
f, g	Known data in the pseudodiagnosticity task
a	Initially provided (anchor) likelihood value
r	Row cell value
c	Column cell value
d	Diagonal cell value
x_a, x_r, x_c, x_d	Continuous variables for the corresponding cells
R, C, D	Information search options in the pseudodiagnosticity task
π	Shorthand for $P(h)$, the initial probability of h in the pseudodiagnosticity task
π_a	Shorthand for $P(h f \wedge g \wedge a)$, the probability of h given the anchor value a along with data f, g

Table 3
A Truth Seeker's Utilities

Chosen hypothesis	True hypothesis	
	<i>h</i>	$\neg h$
<i>h</i>		
Outcome	Hitting <i>h</i>	Missing $\neg h$
Utility	1	0
$\neg h$		
Outcome	Missing <i>h</i>	Hitting $\neg h$
Utility	0	1

tion has not been acquired yet, will be denoted as $EU_x(H)$. It is computed as the expected value of the quantity in Equation 4 across all possible outcomes of the search concerning X , as follows:

$$EU_x(H) = \int_{x|_{h_i \in H}} \max(P(h_i|x))p(x)dx. \tag{5}$$

In the form displayed here, Equation 5 implies X being a continuous set of possible values for variable x (p denotes the relevant probability density function). The expected utility gain associated with an information search about X in view of a choice within H , here denoted as $\Delta EU(H, X)$, amounts to the simple difference between the quantities in Equations 5 and 3, respectively; that is,

$$\Delta EU(H, X) = EU_x(H) - EU(H). \tag{6}$$

In order to show how this model applies to standard problems in pseudodiagnosticity experiments, we will need to have a fresh look at the probabilistic structure of those problems.

Assumptions

We will now state a set of assumptions for our analysis along with some relevant points of notation. Relying on a 2×2 array such as the one instantiated in Table 1, let us define four variables $x_a, x_r, x_c,$ and x_d for the values in the table cells (subscripts stand for anchor, row, column, and diagonal, respectively). Variables $x_a, x_r, x_c,$ and x_d are assumed to be continuous and distributed over $[0, 1]$. Letters $a, r, c,$ and d will denote corresponding values. We will also stipulate to employ the label h for the hypothesis to which the initially provided anchor value a (as well as the column value) refers. The alternative hypothesis (to which both the row and the diagonal value refer) will thus be labeled $\neg h$ (notice that the stipulated notation can be applied to any variant of Table 1, no matter which likelihood is initially provided, that is, where the anchor value is located in the table). The standard version of the pseudodiagnosticity paradigm will be modeled by providing a rational agent facing the experimental task with a set of four basic assumptions, to be stated and motivated below.

As described earlier, participants in the pseudodiagnosticity task are given (or otherwise assumed to entertain) base-rate or prior values for the hypotheses at issue, h and $\neg h$. Importantly, such probabilities are meant to be “prior” to the consideration of data f and g and to be unaffected by what the content of the likelihood table may happen to be. Notice that the expression $P(h|a \wedge r \wedge c \wedge d)$

denotes the probability of h given a fully disclosed likelihood table but with f and g not being known as data. In our current framework, thus, participants are provided with the following assumption about priors:

$$\text{Priors: For whatever } a, r, c \text{ and } d, P(h|a \wedge r \wedge c \wedge d) = P(h).$$

Second, we will assume our agent to rely on the probabilistic independence of f and g , at least conditional on h and $\neg h$, and once again, for any combination of cell values:

$$\text{Data (conditional) independence: For whatever } a, r, c, \text{ and } d,$$

$$P(f \wedge g|h \wedge a \wedge r \wedge c \wedge d) = P(fh \wedge a \wedge r \wedge c \wedge d)$$

$$\times P(gh \wedge a \wedge r \wedge c \wedge d) = ac;$$

and

$$P(f \wedge g|\neg h \wedge a \wedge r \wedge c \wedge d) = P(f\neg h \wedge a \wedge r \wedge c \wedge d)$$

$$\times P(g\neg h \wedge a \wedge r \wedge c \wedge d) = rd.$$

Conditional independence of the given data f and g has been explicitly mentioned as a basic assumption for the usual interpretation of the task (Mynatt et al., 1993, p. 762). Indeed, as already noticed, standard pseudodiagnosticity tasks have typically involved f and g as indicating conditions that appear to be independent in the sense above. This is particularly clear in clinical scenarios, wherein pairs of symptoms such as cough and leg pain or fever and rash have been employed (Kern & Doherty, 1982; Wolf et al., 1985). Such pairs of symptoms represent possible effects of a common cause, that is, an underlying disease (referred to by either hypothesis h or $\neg h$), while apparently bearing no direct physiopathological connection. It is a widely accepted principle that common causes “screen off” their effects, that is, make them conditionally independent (Pearl, 2000; Reichenbach, 1956; Spirtes, Glymour, & Shienes, 1993).

Finally, the two following further conditions will be posited:

Uniformity: Variables $x_a, x_r, x_c,$ and x_d are uniformly distributed.

Likelihood independence: Variables $x_a, x_r, x_c,$ and x_d are independent.

On the whole, the latter assumptions are meant to capture a trait that has been deliberately pursued by researchers in devising variants of the standard task, i.e., the neutralization of the otherwise possible impact of background knowledge on expectations about the cell values. The usual reading of the standard pseudodiagnosticity task has been presented and defended precisely on the basis that “no valid grounds exist for making reasonable *a priori* estimates” about such values (Mynatt et al., 1993, p. 762). Now, it is a celebrated epistemological principle (the *principle of indifference*) advocated in various guises by chief Bayesian theorists such as Keynes (1921), Jeffreys (1939), and Carnap (1962) that a rational agent should assign equal probabilities to two (or more) incompatible events unless reasons are known (or believed) to do otherwise. Given that, the uniformity assumption seems an entirely natural way to model the participants’ basic state of ignorance as set by the standard pseudodiagnosticity task. The likelihood independence assumption, moreover, implies that knowledge of the

initially provided value a (or of any other value, if subsequently discovered) would not, by itself, alter the above state of ignorance as to the cells left undisclosed.

The Epistemic Utility Analysis at Work

Preliminary Results: No Information Search

A crucial tenet of the usual reading of the pseudodiagnosticity task is the claim that updating the priors of h and $\neg h$ by normatively appropriate computations is allowed if and only if the likelihood value in row r is disclosed along with the initially provided likelihood a . Before presenting a full-fledged epistemic utility analysis of the task, then, it will be useful to assess, as a preliminary step, the tenability of this claim. We will now argue that it is actually false.

To begin with, consider a perfectly Bayesian truth-seeking agent who has to face a scenario such as the one illustrated in Table 1. Also suppose that the agent endorses the priors, data independence, uniformity, and likelihood independence assumptions. Finally, suppose that this agent does not have the opportunity to search for any of the missing pieces of information (r , c , or d). She just has to choose h or $\neg h$, with the only information given being the precise value of a single likelihood (a), along with data f and g . How should such a task be carried out?

By assumption, the agent would maximize expected utility by simply choosing the most probable hypothesis. By the usual reading of the pseudodiagnosticity experimental task, however, the agent cannot but make a random choice, the claim being that, even if f and g are known and are potentially relevant data, priors cannot be appropriately updated on the basis of the provided value a alone (i.e., in the absence of any completely specified likelihood ratio). A detailed probabilistic analysis, however, yields a different conclusion.

We are interested in the probabilities of h versus $\neg h$ given data f and g along with knowledge that the anchor likelihood amounts to a certain value a , that is, $P(h|f \wedge g \wedge a)$ and $P(\neg h|f \wedge g \wedge a)$. Now, if x_r , x_c , and x_d are taken as continuous variables over $[0, 1]$ and the priors, data independence, uniformity, and likelihood independence assumptions are made, then the following demonstrably holds (for ease of notation, π is employed to denote the prior probability of h taken as a parameter, that is, not fixed at any specific value):

$$P(h|f \wedge g \wedge a) = \frac{a\pi}{a\pi + \frac{1}{2}(1 - \pi)} \tag{7a}$$

and

$$P(\neg h|f \wedge g \wedge a) = \frac{\frac{1}{2}(1 - \pi)}{a\pi + \frac{1}{2}(1 - \pi)} \tag{7b}$$

The formal derivation of Equations 7a and 7b is carried out in detail in Appendix A. For our present purposes, the crucial point is that, once the value a is revealed (along with knowledge of data f and g), posterior probabilities of h and $\neg h$ can be computed accordingly and will usually depart from the background priors, thus implying a nontrivial updating even with only one table value being provided. In particular, if the additional condition is included that priors are equal (i.e., .50), then one has the following simple functions of a (graphically plotted in Figure 1):

$$P(h|f \wedge g \wedge a) = \frac{a}{a + \frac{1}{2}} \tag{8a}$$

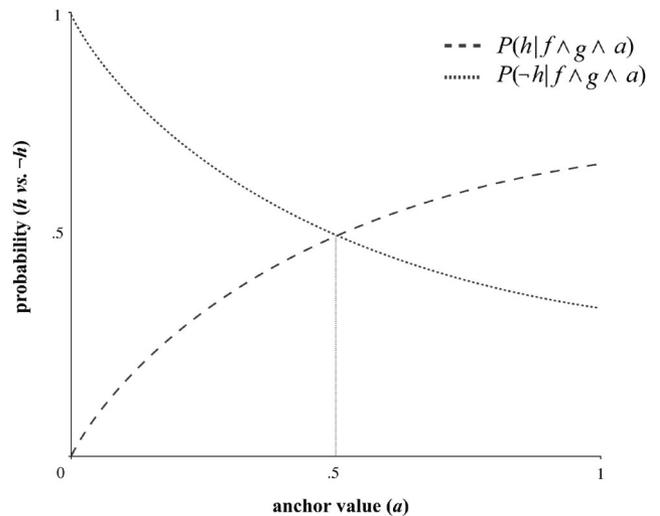


Figure 1. Updated probabilities of h versus $\neg h$ as functions of a , assuming equal priors.

and

$$P(\neg h|f \wedge g \wedge a) = \frac{\frac{1}{2}}{a + \frac{1}{2}} \tag{8b}$$

By a straightforward instantiation of Equation 4 above, the expected epistemic utility of choosing h versus $\neg h$ given a , without any further search for information being pursued, is

$$EU(H|f \wedge g \wedge a) = \max[P(h|f \wedge g \wedge a), P(\neg h|f \wedge g \wedge a)]. \tag{9}$$

As implied by Equations 7a and 7b, $P(h|f \wedge g \wedge a)$ and $P(\neg h|f \wedge g \wedge a)$ can be immediately ranked, depending on the priors and a . A rational agent would then be able to make a principled, nonrandom utility-maximizing choice of h versus $\neg h$ on the basis of the value of likelihood a alone (along with given data f and g). As to the special case of equal priors, Equations 8a and 8b show (and Figure 1 illustrates) that indeed it all depends on whether a is higher or lower than .50. For example, if priors are equal and $a = .65$ (as in Mynatt et al., 1993, Exp. 1), then $P(h|f \wedge g \wedge a) = .65 / (.65 + .50) \approx .57$ and $P(\neg h|f \wedge g \wedge a) = .50 / (.65 + .50) \approx .43$.

Capturing the Standard Pseudodiagnosticity Task

Our next step is facing this question: Given her stated assumptions, how should a rational (perfectly Bayesian) truth-seeking agent search for information in a standard pseudodiagnosticity task? The agent has three alternative strategies available to gain additional information, i.e., searching for the row, the column, or the diagonal value, labeled search strategies R , C , and D , respectively, hereafter. The issue is then to compute the expected utility gain associated with each option, that is,

$$\Delta EU(H, R|f \wedge g \wedge a) = EU_R(H|f \wedge g \wedge a) - EU(H|f \wedge g \wedge a) \tag{10}$$

$$\begin{aligned} \Delta EU(H, C|f \wedge g \wedge a) &= EU_C(H|f \wedge g \wedge a) - EU(H|f \wedge g \wedge a) \quad (11) \end{aligned}$$

$$\begin{aligned} \Delta EU(H, D|f \wedge g \wedge a) &= EU_D(H|f \wedge g \wedge a) - EU(H|f \wedge g \wedge a) \quad (12) \end{aligned}$$

The computation of quantity $EU(H|f \wedge g \wedge a)$ has already been addressed in the above paragraph. Now suppose a row search has been performed and value r has been discovered. Then, by Equation 4 above, we would have

$$\begin{aligned} EU(H|f \wedge g \wedge a \wedge r) &= \max[P(h|f \wedge g \wedge a \wedge r), P(\neg h|f \wedge g \wedge a \wedge r)]. \quad (13) \end{aligned}$$

On the basis of Equation 5 above, in order to have the expected utility of choosing h versus $\neg h$ when an information search about r is going to be pursued but the value of r has not been discovered yet, we have to compute the mean value of $EU(H|f \wedge g \wedge a \wedge r)$ across the possible values of r , that is,

$$\begin{aligned} EU_R(H|f \wedge g \wedge a) &= \int_0^1 \max[P(h|f \wedge g \wedge a \wedge x_r), \\ &P(\neg h|f \wedge g \wedge a \wedge x_r)]P(x_r|f \wedge g \wedge a)dx_r. \quad (14) \end{aligned}$$

For ease of notation, we will now introduce π_a as denoting $P(h|f \wedge g \wedge a)$. Notice that, by the foregoing analysis, this quantity is perfectly defined in any instance of the standard pseudodiagnosticity task as depending on $P(h)$ and the anchor value a (see Equation 7a). By the priors, data independence, uniformity, and likelihood independence assumptions, along with Equations 13 and 14, Equation 10 can be solved as determining the value of $\Delta EU(H, R|f \wedge g \wedge a)$ as an algebraic function of π_a as follows:

$$\begin{aligned} \Delta EU(H, R|f \wedge g \wedge a) &= \begin{cases} \frac{\pi_a^2}{4(1 - \pi_a)} & \text{for } 0 \leq \pi_a \leq 1/2 \\ \frac{(2 - 3\pi_a)^2}{4(1 - \pi_a)} & \text{for } 1/2 < \pi_a \leq 2/3 \\ 0 & \text{for } 2/3 < \pi_a \leq 1. \end{cases} \quad (15) \end{aligned}$$

It may be useful to comment on the meaning of the last row of Equation 15. Suppose that π_a —that is, $P(h|f \wedge g \wedge a)$ —is higher than $2/3$. Then it can be shown that, in order to switch the current choice for the most probable hypothesis h to a subsequent choice for $\neg h$, the actual value r should turn out to be higher than 1, which is of course impossible. (The expression for $P(h|f \wedge g \wedge a \wedge r)$ being employed for this result is displayed in Appendix 1.) So, if $\pi_a > 2/3$, then hypothesis h will still be chosen after a row search, no matter what. For this reason, the expected utility of such a subsequent choice will remain π_a , just as it is with no information search being pursued. In other terms, this is a class of cases in which a row search provides no gain in expected epistemic utility.

A detailed derivation of Equation 15 appears in Appendix B, along with the calculations yielding the corresponding expressions for $\Delta EU(H, C|f \wedge g \wedge a)$ and $\Delta EU(H, D|f \wedge g \wedge a)$, that is,

$$\begin{aligned} \Delta EU(H, C|f \wedge g \wedge a) &= \begin{cases} 0 & \text{for } 0 \leq \pi_a \leq 1/3 \\ \frac{(3\pi_a - 1)^2}{4\pi_a} & \text{for } 1/3 < \pi_a \leq 1/2 \\ \frac{(1 - \pi_a)^2}{4\pi_a} & \text{for } 1/2 < \pi_a \leq 1 \end{cases} \quad (16) \end{aligned}$$

and

$$\begin{aligned} \Delta EU(H, D|f \wedge g \wedge a) &= \begin{cases} \frac{\pi_a^2}{4(1 - \pi_a)} & \text{for } 0 \leq \pi_a \leq 1/2 \\ \frac{(2 - 3\pi_a)^2}{4(1 - \pi_a)} & \text{for } 1/2 < \pi_a \leq 2/3 \\ 0 & \text{for } 2/3 < \pi_a \leq 1. \end{cases} \quad (17) \end{aligned}$$

Once again, null values of expected utility gain reflect classes of cases in which the outcome of the corresponding information search cannot possibly alter current hypothesis choice.

Equations 15–17 represent the gain in expected utility of a search about r , c , or d for a subsequent choice of h versus $\neg h$, as calculated when the actual value of the selected variable is not available yet. They will thus allow a comparison of the usefulness associated with each information search strategy available in standard pseudodiagnosticity experimental procedures.

Optimal Versus Observed Behavior Again

A Theorem

A remarkable fact immediately emerges from Equations 15 and 17: $\Delta EU(H, R|f \wedge g \wedge a)$ and $\Delta EU(H, D|f \wedge g \wedge a)$ amount to identical algebraic expressions. Thus, searching for either the row or the diagonal value in the standard pseudodiagnosticity task provides the same expected utility gain for a truth-seeking rational agent. Recall that strategies R and D are critically opposite for the usual reading: The latter, in contrast to the former, is said to be associated with worthless information and therefore irrational behavior. In our utility-based analysis, on the contrary, R and D turn out to be formally indistinguishable in the standard case. A straightforward consequence is then that, even if it were optimal (which it often is not—see below), search strategy R could not possibly be the only rational option, as has been usually claimed.

We can now come to the core result of the present analysis, that is, the assessment of the usefulness of information search strategy C as compared with R (and, equivalently, with D). How are quantities $\Delta EU(H, C|f \wedge g \wedge a)$ and $\Delta EU(H, R|f \wedge g \wedge a) = \Delta EU(H, D|f \wedge g \wedge a)$ related for different possible values of the parameter π_a as implied by the priors and the anchor? This relation turns out to be governed by the following theorem:

Theorem: In standard pseudodiagnosticity tasks (i.e., under the priors, data independence, uniformity, and likelihood independence assumptions):

for $\pi_a = 0, 1/2, 1$, $\Delta EU(H, C|f \wedge g \wedge a)$
 $= \Delta EU(H, R|f \wedge g \wedge a) = \Delta EU(H, D|f \wedge g \wedge a)$;

for $0 < \pi_a < 1/2$, $\Delta EU(H, C|f \wedge g \wedge a)$
 $< \Delta EU(H, R|f \wedge g \wedge a) = \Delta EU(H, D|f \wedge g \wedge a)$;

and

for $1/2 < \pi_a < 1$, $\Delta EU(H, C|f \wedge g \wedge a)$
 $> \Delta EU(H, R|f \wedge g \wedge a) = \Delta EU(H, D|f \wedge g \wedge a)$.

For a proof, see Appendix C. For an informal suggestion of the underlying analysis, consider the following: Suppose that π_a —that is, $P(h|f \wedge g \wedge a)$ —is moderately high, say .60, so that hypothesis h should be initially chosen. Then there are quite a few low values to be possibly found in the column cell that would alter hypothesis selection with rather dramatic effects. For instance, should it turn out that, say, $c = .10$, then it can be computed that the probability of $\neg h$ would steeply jump up from .40 to about .77 (the equations being employed once again come from Appendix A). On the other hand, comparably high values to be possibly found in the row (or diagonal) cell would still change hypothesis selection, but not with comparably large effects. To illustrate, should it turn out that $r = .90$, then it can be computed that the probability of $\neg h$ would rise from .40 and surpass h but still be just about .55. This reflects a column search being of higher expected value with a high π_a .

An exactly opposite pattern obtains in the symmetric case $\pi_a = .40$. If so, then it can be computed that a .90 high value to be possibly found in the column cell would alter hypothesis selection from an initial choice for $\neg h$ to a subsequent choice for h with a final probability of .55 only, whereas a .10 low value to be possibly found in the row (or diagonal) cell would set the posterior of h to

.77. This reflects a row (or diagonal) search being of higher expected value with a low π_a .

Standard Version With Equal Priors

Our theorem above has, on the basis of Equations 8a and 8b, an immediate corollary concerning standard versions of the pseudodiagnosticity task in which equal priors are assumed.

Corollary: In standard pseudodiagnosticity tasks with equal priors, that is, $P(h) = P(\neg h)$,

for $a = 0, 1/2$, $\Delta EU(H, C|f \wedge g \wedge a)$
 $= \Delta EU(H, R|f \wedge g \wedge a) = \Delta EU(H, D|f \wedge g \wedge a)$;

for $0 < a < 1/2$, $\Delta EU(H, C|f \wedge g \wedge a)$
 $< \Delta EU(H, R|f \wedge g \wedge a) = \Delta EU(H, D|f \wedge g \wedge a)$;

and

for $1/2 < a \leq 1$, $\Delta EU(H, C|f \wedge g \wedge a)$
 $> \Delta EU(H, R|f \wedge g \wedge a) = \Delta EU(H, D|f \wedge g \wedge a)$.

A relevant graphical representation emerges from Figure 2, wherein quantities $\Delta EU(H, C|f \wedge g \wedge a)$ and $\Delta EU(H, R|f \wedge g \wedge a) = \Delta EU(H, D|f \wedge g \wedge a)$ are plotted together for all values of a , corresponding to different possible experimental scenarios.

It can now be seen that the divergence of the present analysis from the commonly adopted reading of the task is stark. First of all, consider the values of the anchor that have usually been employed in standard scenarios with equal priors, from moderately low (such as .35, as in Mynatt et al., 1993, Exp. 2) to high (such as .84, as in Kern & Doherty, 1982). As far as these values are concerned, no search strategy turns out to be completely worthless for a Bayesian truth seeker: Each of them implies at least some expected gain in epistemic utility.

Furthermore, searching for the column value (search strategy C), usually taken as indicating a cognitive bias, is in fact optimal

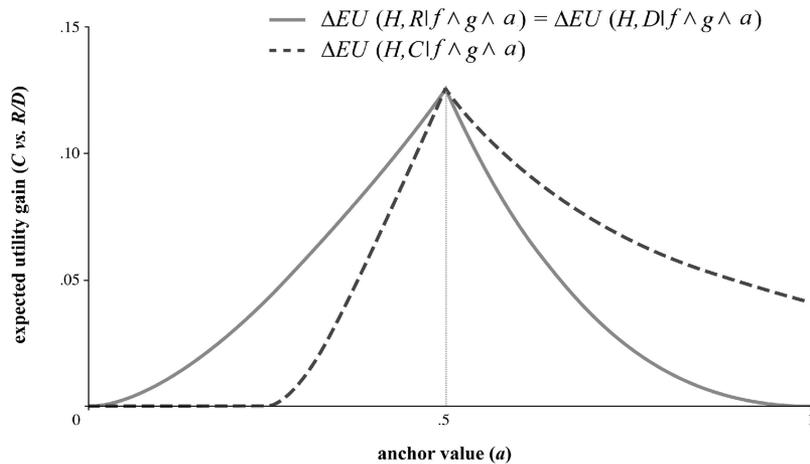


Figure 2. Standard pseudodiagnosticity tasks with equal priors: the expected utility gain of alternative search strategies as functions of a .

provided that the value of a specified in the experimental scenario exceeds the threshold of .50. To illustrate, suppose that $a = .65$, as in Mynatt et al. (1993, Exp. 1); then search strategy C (chosen 59% of the times in this experiment; Mynatt et al., 1993, p. 769) yields an expected utility gain of about .084 versus about .053 as provided by search strategy R or D . This amounts, respectively, to a relative increase of 15% versus less than 10% of the initially expected utility $EU(H|f \wedge g \wedge a)$, which is about .57, corresponding to $P(h|f \wedge g \wedge a)$. Even more remarkable, the convergence between optimal and observed behavior holds for low-anchor experimental scenarios as well. In fact, suppose that $a = .35$, as in Mynatt et al. (1993, Exp. 2); then search strategy R or D (chosen 57% of the times in this experiment; Mynatt et al., 1993, p. 771) yields an expected utility gain of about .072 versus about .034 as provided by search strategy C . This amounts, respectively, to a relative increase of about 12% versus about 6% of the initially expected utility $EU(H|f \wedge g \wedge a)$, which is about .59, corresponding to $P(\neg h|f \wedge g \wedge a)$.

To sum up, the pattern of responses usually obtained in standard pseudodiagnosticity tasks with equal priors precisely matches results from the rational analysis of those very tasks.

Standard Version With Unequal Priors

According to the common reading of the pseudodiagnosticity paradigm, row information search is the only worthwhile option no matter what. As shown above, such a conclusion is unfounded. Yet, to the best of our knowledge, it has never been challenged in published work. Quite the contrary, it has been taken as a suitable and reliable basis to devise experimental scenarios and to analyze, report, and interpret the data obtained. One unfortunate outcome is that several contributions from the literature simply fail to provide sufficient information for appropriate assessment of optimal behavior.

Inquiries with standard pseudodiagnosticity problems with unequal priors are a case in point. Wolf et al. (1985), for instance, employed three different clinical scenarios instantiating the standard pseudodiagnosticity task with anchor values ranging from .15 to .66 and priors ranging from .50 to .67. Importantly, the current analysis implies that optimal choices were different across the three cases, depending on different values of parameter π_a (see below). However, Wolf et al. reported their results in an aggregated fashion by which no conclusion can be drawn concerning response frequencies in each of their problems; thus no sound normative diagnosis of observed behavior is possible.

In a subsequent study, Wolf et al. (1988) did report some disaggregated choice frequencies from early resident physicians facing the same three problems as in the 1985 experiment. For Case 1 ($\pi = .50$, $a = .66$, thus $\pi_a \approx .57$), suboptimal choices R were 43%— $\Delta EU(H, R|f \wedge g \wedge a) = \Delta EU(H, D|f \wedge g \wedge a) = .050$ versus $\Delta EU(H, C|f \wedge g \wedge a) = .082$ —but the exact frequency of C choices (optimal) was not reported. An anomalous pattern of results was obtained in Case 2 ($\pi = .50$, $a = .58$, thus $\pi_a \approx .54$). Here, slightly suboptimal R choices— $\Delta EU(H, R|f \wedge g \wedge a) = \Delta EU(H, D|f \wedge g \wedge a) = .082$ versus $\Delta EU(H, C|f \wedge g \wedge a) = .099$ —were unusually common (58%) for an anchor that was high (although only slightly so). Anyway, the exact frequency of C choices (optimal) was not reported for this problem either. Interestingly, optimal responses were prevalent in the presence of the

manipulation of priors in Case 3 ($\pi = .67$, $a = .15$, thus $\pi_a \approx .38$). Optimal choice R was selected by a majority (62%), to which the frequency of equally optimal D choices should be added— $\Delta EU(H, R|f \wedge g \wedge a) = \Delta EU(H, D|f \wedge g \wedge a) = .056$ versus $\Delta EU(H, C|f \wedge g \wedge a) = .010$ —but the latter is again not reported in the article.

Gruppen, Wolf, and Billi (1991) also employed unequal priors in two out of three problems in their Experiment 2. It should be noticed that for half the participants in this study, all table values were in view from the beginning (full-information condition); only the other half were given a proper pseudodiagnosticity search task, with only one anchor value being provided (partial-information condition). In any event, the anchor values employed, as well as disaggregated choice frequencies, were not reported in the article. Once again, this is presumably due to endorsement of the usual analysis of the task, on the basis of which such information would be irrelevant for the interpretation of results.

In conclusion, only one sufficiently detailed report has been detected in the literature of a standard pseudodiagnosticity task with unequal priors (i.e., Case 3 in Wolf et al., 1988), with no departure from rational behavior being documented in participants' responses.

Nonstandard Versions

Feeney et al. (2000; 2008, Exp. 2) employed interesting non-standard variants of the pseudodiagnosticity task. To grasp the nature of the problems presented to participants, one can usefully rely on Feeney et al.'s (2008) own illustrative example (p. 214). Participants are asked to imagine a friend of theirs having recently bought a new house. It is on either street A (hypothesis h) or street B (hypothesis $\neg h$), they are told, but they just can't remember which one. They do have, however, two data at their disposal: The house has a swimming pool (datum f ; a garden, in a different condition) and a garage (datum g). They also know that 70% of houses on street A have a swimming pool (in our notation, thus, $a = .70$). The usual search task is then assigned.

The intended manipulation carried out in these experiments is on the judged rarity of datum f (the house having a swimming pool vs. a garden), thus on the expectation of a relatively low versus high value in the row cell on the basis of the participants' background knowledge. We would like to argue, however, that once these kinds of pseudodiagnosticity problems are devised, in which resort to background knowledge is relevant and encouraged, a number of additional factors become involved that have not been fully appreciated. As a consequence, crucial elements for rational analysis of the problems employed and assessment of participants' performance end up being left unspecified. Also, participants turn out to have been presented with problems whose solution is daunting for highly sophisticated analysts, let alone for naive reasoners.

The above claims are supported by the following remarks. To begin with, the rare versus common feature manipulation along with the use of real-life contents produces departures from the uniformity assumption, as suggested by Feeney et al.'s (2008) manipulation check. For instance, in the "common" condition from their Experiment 2, Feeney et al. identified average expected row values as varying from .49 to up to .92 across four problem contents, with average expected column/diagonal values concurrently ranging between .56 and .84. To model rational behavior,

one would then have to posit nonuniform distributions for x_r , x_c and x_d matching average expectations from the participants.¹ For example, to model a mild rarity assumption about the row cell, corresponding to an expected value of, say, $1/3$, one might replace a uniform distribution by a suitable beta distribution with parameters $\alpha = 2$, $B = 4$, that is, posit the following probability density function:

$$p(x_r) = \frac{x_r(1 - x_r)^3}{\int_0^1 u(1 - u)^3 du}$$

Along this line, solution of Equations 10–12 would become considerably harder, yet still viable. Notice, however, that such analyses should be carried out separately for each problem employed and separately compared with corresponding observed behavior. This is already at odds with catch-all diagnoses of rationality/irrationality based on the usual reading of pseudodiagnosticity.

To these complications should be added the need to relax all independence assumptions. In the above house example, in fact, having a swimming pool and having a garage will tend to be perceived as positively associated among houses from any street considered, so that the given data f and g are not independent conditional on either h or $-h$ (in violation of the data conditional independence assumption). Also, coming to know that, say, 70% of houses on either street A or B have a swimming pool (i.e., the anchor or row value, respectively) would probably affect expectations of the proportion of houses on the same street having a garage (i.e., the column or diagonal value, respectively), for streets with more houses having a swimming pool will tend to be such that more houses presumably have a garage, so that cell values are not independent (in violation of the likelihood independence assumption). The latter remark is consistent with a further manipulation check reported by Feeney et al. (2008) for their Experiment 1 (see p. 216, Table 1, columns C and F). However, as no well-defined assumptions appear in these studies concerning the kinds of dependencies at work in each scenario, no rational analysis of the tasks can be safely pursued. Also, were such assumptions provided, the solution of the relevant equations (i.e., 10–12 above) would now become a remarkably complex exercise. The concern legitimately arises, then, as to what extent such challenging problems, intriguing as they can be, are effective in investigating people's naive reasoning abilities in a controlled fashion.

Other Studies of Information Search Behavior

Before presenting some conclusions based on our results concerning the pseudodiagnosticity task, we will briefly extend the discussion to other related studies.

A major source of relevant considerations for information search behavior arises from the extensive and diverse inquiry on Wason's selection task (Wason, 1966, 1968), which is still fostering a lively debate. For a long time thought to elicit a basic form of confirmation bias and irrational behavior (see, e.g., Manktelow & Over, 1993; Stein, 1996; Stich, 1990), the task has been reanalyzed through a sophisticated (Bayesian) account of information search akin to the one employed here, by which participants' responses have been said to be not only vindicated but also actually explained as arising from cognitive processes reflecting

rational data selection (Oaksford & Chater, 1994; also see Fitelson, in press, and Nickerson, 1996). The adequacy of such an explanation has not remained unchallenged, however—one relevant objection being precisely that, unlike the present treatment, it does model participants as departing from received instructions (see Evans & Over, 1996a; Laming, 1996; Oaksford & Chater, 1996, 1998, 2003; and Oberauer, Wilhelm, & Rosas Diaz, 1999, for major contributions to this debate; also see Nelson, 2005, an important review of the issue of information search in the study of human cognition, including a host of further relevant references).

The case of Wason's selection task illustrates to what extent concerns about the theoretical framework may affect understanding of empirical results. An additional example is more closely related to the pseudodiagnosticity literature. Covey and Lovie (1998) devised sophisticated computerized stimuli involving 2×2 arrays similar to pseudodiagnosticity scenarios and reported main effects of layout (i.e., of hypotheses being displayed in columns vs. rows) and of question wording (i.e., of the evaluation of only h vs. both h and $-h$ being asked for). However, Covey and Lovie's study did not have an information search task comparable to that in pseudodiagnosticity experiments, for participants in this study could (and all eventually did) uncover all cells. The dependent variables relevant for information search concerned only the cell sequence followed and the time spent in looking at each disclosed cell. As interesting as they are in their own terms, such measurements clearly do not provide a test for optimal information selection, as any pattern followed eventually led to the same (complete) information available for subsequent judgment. Briefly put, participants were not forced to choose which piece(s) of information to collect within a set of alternatives; they could only arrive at the same evidence by different routes.

Beyth-Maron and Fischhoff (1983) investigated people's information search behavior with variants of the following procedure. At issue was the profession of a person on the background information that s/he was present at a party (b) including only university professors (h) and business executives ($-h$), given the additional evidence that s/he is a member of the Bear's Club (e). Participants were asked to classify the following as relevant/irrelevant: (a) the percentage of people at the party who are university professors, that is, $P(h|b)$; (b) the percentage of the Bear's Club members who are at the party, that is, $P(b|e)$; (c) the percentage of university professors at the party who are members of the Bear's Club, that is, $P(e|h \wedge b)$; and (d) the percentage of business executives at the party who are members of the Bear's Club, that is, $P(e|-h \wedge b)$. The items that are actually relevant, from a Bayesian perspective, are (a), (b), and (d). Observed frequencies of relevant responses were high for items (a) and (b) (consistently above 70%) but ranged from 34.5% to 78% for item (d) upon various manipulations, indicating that participants did not always conform to Bayesian principles in this judgment task, as contrasted to choice tasks such as in pseudodiagnosticity, Wason, or Vuma planet experiments (see below).

In another study, Baron, Beattie, and Hershey (1988) suggested that human reasoners actually depart from Baron's (1985) epistemic utility model in some cases in which this seems to be a normatively appro-

¹ In essence, this would amount to an extension of the utility-based framework outlined by Feeney, Evans, and Clibbens (1997) with reference to a class of simplified scenarios akin to nonstandard pseudodiagnosticity tasks.

prate benchmark. The most direct evidence of biased judgment seems to arise from the fictitious medical scenario in their Experiment 4. It included a number of diagnostic tests that were probabilistically informative, yet unable to switch subsequent hypothesis choice and treatment decisions, hence providing no epistemic utility gain. Such tests were nevertheless rated as moderately useful by participants, thus indicating *information bias*. However, participants also assigned consistently higher ratings to diagnostic tests that were both informative and discriminating for subsequent hypothesis choice (i.e., associated to a positive epistemic utility gain).

Finally, findings by Skov and Sherman (1986) suggest interesting rational tendencies in people's information search behavior. Participants were presented with two equally likely hypotheses h and $\neg h$ corresponding to two species of inhabitants of the fictitious planet Vuma, i.e., Gloms and Fizos. They were also given the probability of occurrence of a number of traits (e.g., the creature drinks gasoline) under each hypothesis. The task was to indicate which traits they would ask about to determine whether a novel creature was a Glom or a Fizo. The experimental results revealed that a major determinant of participants' preference for a question concerning a given trait e was the absolute value of the likelihood difference $P(e|h) - P(e|\neg h)$. Notably, given the experimental conditions, relying on this quantity in order to assess the usefulness of a question turns out to be perfectly optimal in terms of expected utility gain (as shown by Nelson, 2005, p. 983).

To sum up, a review of the literature apart from the pseudodiagnosticity paradigm shows that humans' intuitive information search strategies have been sparsely investigated and rather mixed implications have been drawn. On reflection, however, the results obtained seem to leave the normative adequacy of people's behavior as a largely open issue in many respects.

Conclusion

In the standard pseudodiagnosticity paradigm as described above, a Bayesian truth-seeking attitude as well as the priors, data independence, uniformity, and likelihood independence assumptions naturally apply. Thus, on the basis of the foregoing analysis, the participants' prevailing responses, far from showing irrationality, have been typically optimal as documented so far. In particular, the main tendency in standard experimental scenarios with equal priors has been to pursue search strategy C with high anchor values and search strategy R when the anchor value was low—a pattern of responses that precisely matches that of a utility-maximizing, rational agent in the experimental conditions concerned. As a consequence, the diagnosis of the allegedly widespread cognitive bias commonly labeled *pseudodiagnosticity* lacks sound empirical support in observed behavior.

It should be stressed that this conclusion does not rest in any way on postulating divergences between experimenter and participants in terms of their understanding of the problem structure, the task, or the information provided. Quite the contrary, our claim has been that the assumptions employed for the standard pseudodiagnosticity task are in line with the experimenter's intentions. To this extent, our criticism does not amount in any way to a defense of human rationality on the basis of subtleties of the pragmatics of experiments in human cognition.

Our target has been solely the allegedly normative reading that the task has been given. Such a reading crucially rests on the flawed

assumption that in the pseudodiagnosticity paradigm the likelihood value r is the only information that will allow normatively appropriate computations. However, the well-grounded alternative analysis above reveals that it was simply not the case that participants "actively chose irrelevant information and ignored relevant information which was equally easily available" (Doherty et al., 1979, p. 119). This is why the charge that human reasoners do not have a working understanding of diagnosticity (Doherty et al., 1979, p. 112) or fundamentally misunderstand this basic concept (Klayman, 1995, p. 397) is unsupported by the experimental data available.

Let us state in the clearest way that, as we see the issue, such data do not warrant the opposite conclusion either. In our view, then, people's formally optimal choices in this task do not prove them being rational information search agents overall. Also, the normative approach employed here clearly postulates a highly idealized reasoner. We do not mean to have shown that our analysis should be taken as a psychologically realistic representation of human cognitive processes in the pseudodiagnosticity paradigm, let alone in other settings. Thus, our results do not dispel the possibility that most participants handle the task by relying on some cognitively inexpensive heuristic strategy, such as simply looking for information concerning the hypothesis that is presently more likely (see Evans's, 2007, discussion of *fundamental analytical bias* in this respect). Our results do prove, however, that such postulated heuristics, if present, have yielded outcomes that largely match the normatively justified solutions of standard pseudodiagnosticity problems, thus producing no systematically biased, suboptimal, or otherwise irrational behavior.

More generally, we urge that the usual normative reading of the pseudodiagnosticity task demonstrably provides a deceptively simple picture of the problem. Accordingly, uncritical adoption of such a reading has been shown to have negatively affected empirical inquiry in terms of experimental design, methods, results reporting, and interpretation. To this extent, we also see the current contribution as illustrating a case study in the relevance of appropriate normative references in psychological research on human reasoning and behavior.

In conclusion, the present work is not meant to directly address the issue of human rationality in its generality. The intended upshot is more limited but, in our view, more stringent. It can be summarized as follows: Investigating people's ability and limitations in assessing the diagnosticity of information is certainly of primary interest for the study of reasoning and cognition. However, experimental procedures will have to be devised and understood on the basis of a careful consideration of the nature of information search problems. In our opinion, a thorough rational analysis such as the one presented here may thus be of value for future empirical studies that are much needed. On close inspection, in fact, the evidence obtained from the standard pseudodiagnosticity task, commonly taken for 30 years as establishing the existence of a cognitive bias, has failed to discriminate between normative and nonnormative models of information search behavior. As such, one might well say that the evidence concerned turns out to have been nondiagnostic in that respect.

References

- Baron, J. (1985). *Rationality and intelligence*. New York, NY: Cambridge University Press.
- Baron, J. (2000). *Thinking and deciding*. New York, NY: Cambridge University Press.

- Baron, J., Beattie, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty. *Organizational Behavior and Human Decision Processes*, 42, 88–110.
- Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology*, 45, 1185–1195.
- Carnap, R. (1962). *Logical foundations of probability* (2nd ed.). Chicago, IL: University of Chicago Press.
- Covey, J. A., & Lovie, A. D. (1998). Information selection and utilization in hypothesis-testing: A comparison of process-tracing and structural analysis techniques. *Organizational Behavior and Human Decision Processes*, 75, 56–74.
- Crupi, V., Tentori, K., & Gonzalez, M. (2007). On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science*, 74, 229–252.
- Dawes, R. M. (2001). *Everyday irrationality*. Cambridge, MA: Westview.
- 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 & Cognition*, 24, 644–654.
- Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica*, 43, 111–121.
- Evans, J. St. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Hove, United Kingdom: Psychology Press.
- Evans, J. St. B. T., & Over, D. E. (1996a). Rationality in the selection task: Epistemic utility versus uncertainty reduction. *Psychological Review*, 103, 356–363.
- Evans, J. St. B. T., & Over, D. E. (1996b). *Reasoning and rationality*. Hove, United Kingdom: Erlbaum.
- Evans, J. St. B. T., Venn, S., & Feeney, A. (2002). Implicit and explicit processes in a hypothesis testing task. *British Journal of Psychology*, 93, 31–46.
- Feeney, A., Evans, J. St. B. T., & Clibbens, J. (1997). Probabilities, utilities and hypothesis testing. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th Annual Conference of the Cognitive Science Society* (pp. 217–222). Mahwah, NJ: Erlbaum.
- Feeney, A., Evans, J. St. B. T., & Venn, S. (2000). A rarity heuristic for hypothesis testing. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 119–124). Mahwah, NJ: Erlbaum.
- Feeney, A., Evans, J. St. B. T., & Venn, S. (2008). Rarity, pseudodiagnosticity and Bayesian reasoning. *Thinking and Reasoning*, 14, 209–230.
- Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, 90, 239–260.
- Fitelson, B. (2001). A Bayesian account of independent evidence with applications. *Philosophy of Science*, 68, S123–140.
- Fitelson, B. (in press). Bayesian confirmation theory and the Wason selection task. *Synthese*.
- Good, I. J. (1950). *Probability and the weight of evidence*. London, England: Griffin.
- Gruppen, L. D., Wolf, F. M., & Billi, J. E. (1991). Information gathering and integration as sources of error in diagnostic decision making. *Medical Decision Making*, 11, 233–239.
- Jeffreys, H. (1939). *Theory of probability*. Oxford, United Kingdom: Oxford University Press.
- Kern, L., & Doherty, M. E. (1982). "Pseudodiagnosticity" in an idealized medical problem-solving environment. *Journal of Medical Education*, 57, 100–104.
- Keynes, J. M. (1921). *A treatise of probability*. London, England: Macmillan.
- Klayman, J. (1995). Varieties of confirmation bias. In J. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Decision making from a cognitive perspective* (pp. 365–418). New York, NY: Academic Press.
- Laming, D. (1996). On the analysis of irrational data selection: A critique of Oaksford and Chater (1994). *Psychological Review*, 103, 364–373.
- Manktelow, K. I. (1999). *Reasoning and thinking*. Hove, United Kingdom: Taylor & Francis.
- Manktelow, K. I., & Over, D. E. (1993). *Rationality: Psychological and philosophical perspectives*. London, England: Routledge.
- 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.
- Nelson, J. D. (2005). Finding useful questions: On Bayesian diagnosticity, probability, impact and information gain. *Psychological Review*, 112, 979–999.
- Nickerson, R. S. (1996). Hempel's paradox and Wason's selection task: Logical and psychological puzzles of confirmation. *Thinking and Reasoning*, 2, 1–31.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175–220.
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608–631.
- Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review*, 103, 381–391.
- Oaksford, M., & Chater, N. (1998). A revised rational analysis of the selection task: Exceptions and sequential sampling. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 372–398). Oxford, United Kingdom: Oxford University Press.
- Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review, and reevaluation. *Psychonomic Bulletin & Review*, 10, 289–318.
- Oberauer, K., Wilhelm, O., & Rosas Diaz, R. (1999). Bayesian rationality for the Wason selection task? A test of optimal data selection theory. *Thinking and Reasoning*, 5, 115–144.
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, MA: Cambridge University Press.
- Reichenbach, H. (1956). *The direction of time*. Berkeley, CA: University of Los Angeles Press.
- Sackett, D. L., Haynes, R. B., & Tugwell, P. (1985). *Clinical epidemiology*. Boston, MA: Little & Brown.
- Savage, L. J. (1954). *The foundations of statistics*. New York, NY: Wiley.
- Skov, R. B., & Sherman, S. J. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology*, 22, 93–121.
- Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction and search*. New York, NY: Springer.
- Stein, E. (1996). *Without good reason*. Oxford, United Kingdom: Oxford University Press.
- Stich, S. (1990). *The fragmentation of reason*. Cambridge, MA: MIT Press.
- Wason, P. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmondsworth, United Kingdom: Penguin.
- Wason, P. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20, 273–281.
- Wolf, F. M., Gruppen, L. D., & Billi, J. E. (1985). Differential diagnosis and the competing-hypotheses heuristic. *Journal of the American Medical Association*, 253, 2858–2862.
- Wolf, F. M., Gruppen, L. D., & Billi, J. E. (1988). Use of the competing-hypotheses heuristic to reduce pseudodiagnosticity. *Journal of Medical Education*, 63, 548–554.
- Wood, B. P. (1999). Decision making in radiology. *Radiology*, 211, 601–603.

Appendix A

Derivation of Equations 7a and 7b and Related Calculations

To compute $P(hlf \wedge g \wedge a)$, let us first notice that²

$$P(hlf \wedge g \wedge a) = \frac{P(h \wedge f \wedge g \wedge a)}{P(f \wedge g \wedge a)}$$

$$= \frac{P(h \wedge f \wedge g \wedge a)}{P(h \wedge f \wedge g \wedge a) + P(\neg h \wedge f \wedge g \wedge a)}$$

Now,

$$P(h \wedge f \wedge g \wedge a)$$

$$= \int_0^1 \int_0^1 \int_0^1 P(h \wedge f \wedge g \wedge a \wedge x_r \wedge x_c \wedge x_d) dx_r dx_c dx_d$$

$$= \int_0^1 \int_0^1 \int_0^1 P(f \wedge g | h \wedge a \wedge x_r \wedge x_c \wedge x_d)$$

$$\times P(h | a \wedge x_r \wedge x_c \wedge x_d) p(a \wedge x_r \wedge x_c \wedge x_d) dx_r dx_c dx_d.$$

By the priors, data independence, and likelihood independence assumptions,

$$P(h \wedge f \wedge g \wedge a) = \int_0^1 \int_0^1 \int_0^1 ax_c P(h) p(a) p(x_r) p(x_c) p(x_d) dx_r dx_c dx_d$$

$$= a \times p(a) P(h) \int_0^1 x_c p(x_c) dx_c = \overline{ax_c} p(a) P(h).$$

(Variable names with an upperscore denote their respective mean value.)

By a similar computation,

$$P(\neg h \wedge f \wedge g \wedge a) = \overline{x_r} \overline{x_d} p(a) P(\neg h).$$

Thus,

$$P(hlf \wedge g \wedge a) = \frac{\overline{ax_c} p(a) P(h)}{\overline{ax_c} p(a) P(h) + \overline{x_r x_d} p(a) P(\neg h)}$$

$$= \frac{\overline{ax_c} P(h)}{\overline{ax_c} P(h) + \overline{x_r x_d} P(\neg h)}.$$

Finally, by the uniformity assumption, Equation 7a follows with $P(h) = \pi$:

$$P(hlf \wedge g \wedge a) = \frac{a\pi}{a\pi + 1/2(1 - \pi)},$$

from which, by the probability calculus, one immediately has Equation 7b:

$$P(\neg hlf \wedge g \wedge a) = \frac{1/2(1 - \pi)}{a\pi + 1/2(1 - \pi)}.$$

Similar calculations yield the following results:

$$P(hlf \wedge g \wedge a \wedge r) = \frac{a\pi}{a\pi + r(1 - \pi)} = \frac{1/2\pi_a}{1/2\pi_a + r(1 - \pi_a)},$$

$$P(hlf \wedge g \wedge a \wedge c) = \frac{ac\pi}{ac\pi + 1/4(1 - \pi)} = \frac{c\pi_a}{c\pi_a + 1/2(1 - \pi_a)},$$

and

$$P(hlf \wedge g \wedge a \wedge d) = \frac{a\pi}{a\pi + d(1 - \pi)} = \frac{1/2\pi_a}{1/2\pi_a + d(1 - \pi_a)},$$

where $\pi_a = P(hlf \wedge g \wedge a)$.

Appendix B

Derivation of Equations 15–17

Let us first recall Equation 14:

$$= \int_0^1 \max[P(h \wedge x_r | f \wedge g \wedge a), P(\neg h \wedge x_r | f \wedge g \wedge a)] dx_r.$$

$$EU_R(Hlf \wedge g \wedge a)$$

$$= \int \max[P(hlf \wedge g \wedge a \wedge x_r), P(\neg hlf \wedge g \wedge a \wedge x_r)]$$

$$\times P(x_r | f \wedge g \wedge a) dx_r,$$

² Throughout the appendixes, it is assumed that (a) either $a \neq 0$ or $r \neq 0$ and (b) either $c \neq 0$ or $d \neq 0$. Although inconsequential for all quantitative results presented, such assumptions are required to ensure mathematical coherence.

(Appendixes continue)

It can then be shown that, by the priors, data independence, uniformity, and likelihood independence assumptions,

$$P(h \wedge x_r | f \wedge g \wedge a) = P(h | f \wedge g \wedge a) = \pi_a$$

and

$$P(-h \wedge x_r | f \wedge g \wedge a) = 2x_r P(-h | f \wedge g \wedge a) = 2x_r(1 - \pi_a).$$

Thus,

$$EU_R(H | f \wedge g \wedge a) = \int_0^1 \max[\pi_a, 2x_r(1 - \pi_a)] dx_r.$$

Parallel derivations yield

$$EU_C(H | f \wedge g \wedge a) = \int_0^1 \max[2\pi_a x_c, (1 - \pi_a)] dx_c$$

and

$$EU_D(H | f \wedge g \wedge a) = \int_0^1 \max[\pi_a, 2x_d(1 - \pi_a)] dx_d.$$

Now consider $EU_R(H | f \wedge g \wedge a)$. Notice that

$$\pi_a \geq 2x_r(1 - \pi_a) \quad \text{iff} \quad x_r \leq \frac{1}{2} \frac{\pi_a}{(1 - \pi_a)}.$$

If $\pi_a \geq 2/3$, the latter inequality holds for any possible value of x_r (since $x_r \leq 1$). Thus,

$$\text{for } 2/3 \leq \pi_a \leq 1, \quad EU_R(H | f \wedge g \wedge a) = \int_0^1 \pi_a dx_r = \pi_a.$$

On the other hand,

for $0 \leq \pi_a < 2/3$, $EU_R(H | f \wedge g \wedge a)$

$$\begin{aligned} &= \int_0^{1/2 \frac{\pi_a}{(1 - \pi_a)}} \pi_a dx_r + \int_{1/2 \frac{\pi_a}{(1 - \pi_a)}}^1 2(1 - \pi_a)x_r dx_r \\ &= \pi_a \int_0^{1/2 \frac{\pi_a}{(1 - \pi_a)}} dx_r + 2(1 - \pi_a) \int_{1/2 \frac{\pi_a}{(1 - \pi_a)}}^1 x_r dx_r \\ &= \pi_a \left(\frac{1}{2} \frac{\pi_a}{(1 - \pi_a)} \right) + (1 - \pi_a) \left(1 - \left(\frac{1}{2} \frac{\pi_a}{(1 - \pi_a)} \right)^2 \right). \end{aligned}$$

The following equalities are then obtained by algebraic manipulations:

For $0 \leq \pi_a < 1/2$, $\Delta EU(H, R | f \wedge g \wedge a)$

$$\begin{aligned} &= \pi_a \left(\frac{1}{2} \frac{\pi_a}{(1 - \pi_a)} \right) + (1 - \pi_a) \left(1 - \left(\frac{1}{2} \frac{\pi_a}{(1 - \pi_a)} \right)^2 \right) \\ &\quad - (1 - \pi_a) = \frac{\pi_a^2}{4(1 - \pi_a)}. \end{aligned}$$

For $1/2 \leq \pi_a < 2/3$, $\Delta EU(H, R | f \wedge g \wedge a)$

$$\begin{aligned} &= \pi_a \left(\frac{1}{2} \frac{\pi_a}{(1 - \pi_a)} \right) + (1 - \pi_a) \left(1 - \left(\frac{1}{2} \frac{\pi_a}{(1 - \pi_a)} \right)^2 \right) - \pi_a \\ &= \frac{(2 - 3\pi_a)^2}{4(1 - \pi_a)}. \end{aligned}$$

For $2/3 \leq \pi_a \leq 1$, $\Delta EU(H, R | f \wedge g \wedge a) = \pi_a - \pi_a = 0$.

This concludes the derivation of Equation 15. Calculations concerning $EU_D(H | f \wedge g \wedge a)$ and Equation 17 are identical. It remains to compute $EU_C(H | f \wedge g \wedge a)$. Notice that

$$2\pi_a x_c \leq (1 - \pi_a) \quad \text{iff} \quad x_c \leq \frac{1}{2} \frac{(1 - \pi_a)}{\pi_a}.$$

If $\pi_a < 1/3$, the latter inequality holds for any possible value of x_c (since $x_c \leq 1$). Thus,

for $0 \leq \pi_a < 1/3$, $EU_C(H | f \wedge g \wedge a)$

$$= \int_0^1 (1 - \pi_a) dx_c = (1 - \pi_a).$$

On the other hand,

for $1/3 \leq \pi_a \leq 1$, $EU_C(H | f \wedge g \wedge a)$

$$\begin{aligned} &= \int_0^{1/2 \frac{(1 - \pi_a)}{\pi_a}} (1 - \pi_a) dx_c + \int_{1/2 \frac{(1 - \pi_a)}{\pi_a}}^1 2\pi_a x_c dx_c \\ &= (1 - \pi_a) \int_0^{1/2 \frac{(1 - \pi_a)}{\pi_a}} dx_c + 2\pi_a \int_{1/2 \frac{(1 - \pi_a)}{\pi_a}}^1 x_c dx_c \\ &= (1 - \pi_a) \left(\frac{1}{2} \frac{(1 - \pi_a)}{\pi_a} \right) + \pi_a \left(1 - \left(\frac{1}{2} \frac{(1 - \pi_a)}{\pi_a} \right)^2 \right). \end{aligned}$$

The following equalities are then obtained by algebraic manipulations:

For $0 \leq \pi_a < 1/3$, $\Delta EU(H, C | f \wedge g \wedge a)$

$$= (1 - \pi_a) - (1 - \pi_a) = 0.$$

For $1/3 \leq \pi_a < 1/2$, $\Delta EU(H, C | f \wedge g \wedge a)$

$$\begin{aligned} &= (1 - \pi_a) \left(\frac{1}{2} \frac{(1 - \pi_a)}{\pi_a} \right) + \pi_a \left(1 - \left(\frac{1}{2} \frac{(1 - \pi_a)}{\pi_a} \right)^2 \right) \\ &\quad - (1 - \pi_a) = \frac{(3\pi_a - 1)^2}{4\pi_a}. \end{aligned}$$

For $1/2 \leq \pi_a \leq 1$, $\Delta EU(H, C | f \wedge g \wedge a)$

$$\begin{aligned} &= (1 - \pi_a) \left(\frac{1}{2} \frac{(1 - \pi_a)}{\pi_a} \right) + \pi_a \left(1 - \left(\frac{1}{2} \frac{(1 - \pi_a)}{\pi_a} \right)^2 \right) - \pi_a \\ &= \frac{(1 - \pi_a)^2}{4\pi_a}. \end{aligned}$$

This concludes the derivation of Equation 16.

Appendix C

Proof of the Theorem

Theorem: In standard pseudodiagnosticity tasks (i.e., under the priors, data independence, uniformity, and likelihood independence assumptions),

for $\pi_a = 0, 1/2, 1, \Delta EU(H, C|f \wedge g \wedge a)$

$$= \Delta EU(H, R|f \wedge g \wedge a) = \Delta EU(H, D|f \wedge g \wedge a);$$

for $0 < \pi_a < 1/2, \Delta EU(H, C|f \wedge g \wedge a)$

$$< \Delta EU(H, R|f \wedge g \wedge a) = \Delta EU(H, D|f \wedge g \wedge a);$$

and

for $1/2 < \pi_a < 1, \Delta EU(H, C|f \wedge g \wedge a)$

$$> \Delta EU(H, R|f \wedge g \wedge a) = \Delta EU(H, D|f \wedge g \wedge a).$$

Proof: First, posit $\Delta EU(H, C|f \wedge g \wedge a) = \Delta EU(H, R|f \wedge g \wedge a)$ for $0 \leq \pi_a < 1/3$, that is,

$$0 = \frac{\pi_a^2}{4(1 - \pi_a)}.$$

$\pi_a = 0$ is the only solution, whereas the right side $\Delta EU(H, R|f \wedge g \wedge a)$ is clearly higher for any other value in the interval considered.

Then posit $\Delta EU(H, C|f \wedge g \wedge a) = \Delta EU(H, R|f \wedge g \wedge a)$ for $1/3 \leq \pi_a < 1/2$, that is,

$$\frac{(3\pi_a - 1)^2}{4\pi_a} = \frac{\pi_a^2}{4(1 - \pi_a)}.$$

Algebraic manipulations yield

$$(\pi_a - 1/2)(10(\pi_a - 1/2)^2 - 1/2) = 0,$$

which has no solution in the interval considered. If two continuous functions do not intersect in a given interval, then one must be higher in the whole interval. Then, as $\Delta EU(H, C|f \wedge g \wedge a) < \Delta EU(H, R|f \wedge g \wedge a)$ in $0 < \pi_a < 1/3$, this extends to $0 < \pi_a < 1/2$, for both functions are continuous and no point of intersection exists.

Now posit $\Delta EU(H, C|f \wedge g \wedge a) = \Delta EU(H, R|f \wedge g \wedge a)$ for $2/3 \leq \pi_a \leq 1$, that is,

$$\frac{(1 - \pi_a)^2}{4\pi_a} = 0.$$

$\pi_a = 1$ is the only solution, whereas the left side $\Delta EU(H, C|f \wedge g \wedge a)$ is clearly higher for any other value in the interval considered.

Finally, posit $\Delta EU(H, C|f \wedge g \wedge a) = \Delta EU(H, R|f \wedge g \wedge a)$ for $1/2 \leq \pi_a < 2/3$, that is,

$$\frac{(2 - 3\pi_a)^2}{4(1 - \pi_a)} = \frac{(1 - \pi_a)^2}{4\pi_a}.$$

Algebraic manipulations once again yield

$$(\pi_a - 1/2)(10(\pi_a - 1/2)^2 - 1/2) = 0$$

with $\pi_a = 1/2$ as the only solution in the interval considered. As $\Delta EU(H, C|f \wedge g \wedge a) > \Delta EU(H, R|f \wedge g \wedge a)$ in $2/3 \leq \pi_a < 1$, this extends to $1/2 < \pi_a < 1$, for both functions are continuous and no point of intersection exists. This completes the proof of the theorem.

Received September 19, 2008
 Revision received June 26, 2009
 Accepted June 27, 2009 ■