The Quarterly Journal of Experimental Psychology

ISSN: 0272-4987 (Print) 1464-0740 (Online) Journal homepage: www.tandfonline.com/journals/pqja20

Heuristics in medical and non-medical decision-making

Rachael F. Heller, Herbert D. Saltzstein & William B. Caspe

To cite this article: Rachael F. Heller, Herbert D. Saltzstein & William B. Caspe (1992) Heuristics in medical and non-medical decision-making, The Quarterly Journal of Experimental Psychology, 44:2, 211-235, DOI: 10.1080/02724989243000019

To link to this article: https://doi.org/10.1080/02724989243000019

Published online: 29 May 2007.

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 1992, 44A (2) 211–235

Heuristics in Medical and Non-medical Decision-making

Rachael F. Heller and Herbert D. Saltzstein
Graduate School of the City University of New York, New York, U.S.A.

William B. Caspe
The Bronx–Lebanon Hospital and Medical Center, Bronx, NY, U.S.A.

Pairs of hypothetical medical and non-medical problems were given to 44 pediatric residents at three levels of hospital training. Each problem was designed to detect a specific heuristic-based bias in making diagnoses. Discounting, disregarding base rate, and over-confidence in contextually embedded redundant information were more evident on medical than on non-medical problems. In particular, a greater number of third-year residents disregarded base-rate information than did first- and second-year residents on medical but not on non-medical problems. On medical problems, a greater number of first-year residents expressed greater confidence in redundant information that was contextually embedded than in information that was presented in a listed format. Over one-third of the residents confused prospective and retrospective probabilities; three-fourths showed evidence of augmentation; virtually all residents expressed greater confidence in a diagnosis based on redundant rather than on non-redundant listed information. These latter effects were consistent across training level and occurred on both medical and non-medical problems. The results are discussed in terms of prototype theory and the nature of medical training.

Requests for reprints should be sent to Dr. Herbert D. Saltzstein or Dr. Rachael F. Heller, Psychology, Graduate Center of CUNY, 33 West 42 Street, New York, NY 10036; or Dr. William B. Caspe, Head of Pediatrics, The Bronx–Lebanon Hospital and Medical Center, 1276 Fulton Avenue, Bronx, NY 10456.

Portions of this article are based on a doctoral dissertation by the first author under the supervision of the second author. The research was supported by a PSC-CUNY Research Award to the second author.

The authors wish to thank Dr. Kurt Hirschhorn, Head of Pediatrics at Mt. Sinai Hospital and Medical Center specifically for advice on construction of the medical vignettes and generally for advice in carrying out the research, and to Drs. Howard Ehrlichman, Richard F. Heller, and David O'Brien for critically reading an earlier version of this paper, to Dr. David Rindskopf for advice in statistical analysis, to Danny Choriki for help in computer analysis of the data, and to Drs. Kay Deaux, Florence Denmark, Suzanne Kobasa, and Delmos Jones for advice and criticism as members of Dr. Heller's dissertation committee.

In informal conversations physicians often claim that medical diagnosis is an "art" and therefore cannot be studied scientifically. As it stands, this assertion stands as an obstacle rather than a bridge to understanding the diagnostic process. It is possible, however, to view the issue in a form that encourages rather than discourages further inquiry. To this end, we make the following two assumptions:

1. Medical diagnosis, as generally practised, is neither accomplished by an algorithm, such as some application of Bayes' theorem, nor by reference to a dictionary consisting of sets of symptoms paired with the diseases they presumably reflect; and,

2. The very complexity of the task requires short-cuts in reasoning, usually termed heuristics (e.g. Kahneman & Tversky, 1973), which may be within or outside the awareness of the decision-maker.

The ubiquity of heuristics and heuristic-based biases have been demonstrated in the research literature (e.g. Kahneman, Slovic, & Tversky, 1982; Sherman & Corty, 1984). However, it also has been noted that heuristics, such as representativeness (as manifested in ignoring base-rate information) are not always used, or if used, do not always lead to errors (e.g. Ajzen, 1977; Kruglanski, Friedland, & Farkah, 1984; Zukier & Pepitone, 1984). For explanations based on heuristics to be testable and useful, it is necessary to stipulate under what conditions heuristics will be used and by whom. One strategy for doing so is to examine their use and biases resulting from their use by decision makers at different stages of training. This is the framework for the study reported here.

The contrast made here is not the customary one between experts and non-experts, as in a number of recent studies (e.g., Chi, Glaser, & Farr, 1988; Johnson et al., 1981; Joseph & Patel, 1990; Murphy & Wright, 1984; Patel & Groen, 1986). Rather, comparison is between those relatively new to professional training and those relatively socialized into a profession but not necessarily experts. In many previous studies the assumption is often made that experts are "better" at least in their product or outcome than novices. The aim of such research is to discover how they are better. In eschewing the label "expert", our research does not define those more advanced in training as "better". Rather, we compare decision-makers who have received more training with those who have received less training. We may assume that the former have more domain-specific knowledge (e.g. Joseph & Patel, 1990). Whether or not they are better diagnosticians is an open question. However, one assumption that we do share with researchers on expert–novice differences is that whatever differences we find will be domain-specific (e.g. Joseph & Patel, 1990). The purpose of the research is to determine how those with more advanced training approach diagnostic problems within their own domain differently from those with less advanced training.

Thus, the focus of the study and our approach, in general, is not on outcome or accuracy, but on process. In this connection, we do not assume that reliance on heuristics is equivalent to committing errors. Quite on the contrary, the assumption is made that heuristics are used because they do facilitate the decision-making process without frequent error. However, this, like many studies of heuristics, is designed to demonstrate that the heuristics will be used even in circumstances where they may result in bias or even error. Indeed, our reasoning-within-a-heuristics framework suggests that they may make more errors in their reasoning process as a result of their training. To understand why this might be the case, it is first necessary to consider the nature of that professional training along with the characteristics of heuristics. Before continuing our argument, it is appropriate that we briefly introduce the heuristics under study.

The first is *representativeness* or the strategy for judging the likelihood of an event's belonging to a class of events according to the apparent similarity between the event and that class of events without regard to other information, such as base rate or the overall frequency of the class (Kahneman & Tversky, 1973).

The second and third heuristics come from Kelley's causal attribution theory. They are *discounting* and *augmentation*. Discounting is a strategy that terminates the search for alternative causes of an event once a plausible sufficient cause has been identified. Augmentation is a strategy based on the assumption that the occurrence of an event in the presence of a strong impeding factor necessarily implies the presence of a strong facilitating cause. These are described in greater detail below.

Finally, although it is not really a heuristic nor necessarily the result of a heuristic, confusion of prospective and retrospective probabilities was chosen for study because of its relevance to medical decision-making and because previous research (Eddy, 1982) suggests that it may be responsible for important biases in medical decision-making.

Traditional medical training emphasizes the learning of the typical characteristics of diseases or other medical conditions (e.g. pregnancy, puberty), or *symptomatology*. In current cognitive terms, they learn the *prototypes* of their profession (e.g. Cantor & Mischel, 1977; Rosch & Mervis, 1975). Prototypes in medicine are those "fuzzy" sets of clinical syndromes characterized by typical symptoms and other (e.g. etiological) signs (e.g. Cantor, Smith, French, & Mezzich, 1980). As pointed out by Sherman and Corty (1984) in their review of the literature, prototypes and representativeness are closely linked, and therefore the strengthening of prototypes in a domain of knowledge (such as medicine) should lead to a heavy reliance on representativeness. Reliance on this heuristic should lead to two kinds of biases in decision-making. The first is ignoring base-rate information, often termed the base-rate fallacy (e.g. Ajzen, 1977; Kahneman & Tversky, 1973;

Tversky & Kahneman, 1982; Zukier & Pepitone, 1984). This is because, according to Bayes' theorem, two kinds of information—the prospective probability or diagnostic value of the sign, symptom, or test and the base rate or overall incidence of the condition—are relevant to the probability that an event will occur. Representativeness leads the reasoner to overemphasize the importance of the former and ignore or underemphasize the latter. The second resulting bias is an overreliance on redundant information (e.g. Eddy, 1982). Here, we assume that redundancy enhances the representativeness of the sample information, and therefore increases the judged likelihood that the sample belongs to the population, and/or the confidence in the prediction.

Another part of medical training is the learning of the causes of these diseases or conditions, or their *etiology*. These developing sets of beliefs should lead to an increased reliance on *discounting* and *augmentation*, as described by Kelley (1972) and Ross (1981)—that is, the stronger one's beliefs in a set of causal propositions, the more likely one is to stop the search for alternative causes in the presence of such a plausible cause (discounting) and the invocation of stronger facilitating causes in the presence of an impeding cause (augmentation). The argument is relevant to medical diagnosis for two reasons: (1) Diagnosis is made in part on the basis of whether known etiological factors are present—for example, one would not be likely to diagnose malaria in the continental United States if the patient had not recently travelled to an area where the anopheline mosquito was found. (2) We assume that diagnostic categories may be treated as causes and symptoms as effects, so that making a diagnosis is treated as developing a causal proposition. The above reasoning should lead us to expect an increase in reliance on heuristics (along with the consequent problem-solving biases) with increased training at least up to a certain level of training, and this was our general expectation. Nonetheless, given the lack of empirical evidence of such effects and the uncertainty of knowing how post-graduate training, as distinguished from formal medical education, affects cognitive functioning, the hypotheses are offered as tentative.

Though the confusion between prospective and retrospective probabilities does not fall under any of the above heuristics, and indeed hardly qualifies as a heuristic at all, we have included it for study. We have done so because of Eddy's (1982) extremely interesting study of errors by experienced physicians in prescribing mammography. He found that these physicians frequently prescribed mammograms even when these test results provided redundant information. He interpreted this finding as being due to their confusion between these two kinds of conditional probabilities.

Thus, the basic purpose of the study is to study the reliance on heuristics in medical as opposed to non-medical diagnostic problem-solving. Some familiarity of the reader with the heuristics under study may be assumed, but it is possible that their applicability to medical diagnosis may not be so familiar.

Therefore, what follows is a brief summary of the heuristics as applied to the domain of medical diagnosis. The actual items used appear in Appendix A.

Discounting occurs when the sufficiency of one cause to account for an effect leads the reasoner to discount the possibility of other causes. If we take symptoms to be effects and diagnoses as referring to causes, discounting would lead the diagnostician to settle on a single diagnosis and discount the possibility of alternative diagnoses and especially simultaneous multiple diagnoses. Indeed, what is here presented as a heuristic possibly leading to error is often offered as a principle of diagnosis in medical training such as in the dictum, "one disease for one patient, not one disease for each symptom".

Augmentation occurs when especially strong facilitating causes are required in order to account for an event in the presence of an inhibitory cause. Applied to medical diagnosis, let us consider a counter-indicator as an inhibitory cause—for example, maleness in diagnosing anorexia nervosa, being Caucasian in diagnosing sickle-cell anemia, and having grown up in a tropical climate in diagnosing multiple sclerosis. Hence, an especially strong facilitating cause, here disease, is required to account for the effect, symptom(s). (Underlying augmentation is the assumption of compensating causes, i.e. negative [inhibitory] causes require strong positive [facilitating] causes if the effect has occurred.) With a slight extrapolation of the augmentation heuristic, we are led to predict that the diagnostician requires more stringent clinical indicators of the disease (e.g. weight loss) in order to diagnose it (e.g. anorexia nervosa) in the presence of the counter-indicator (e.g. for a male rather than a female patient), although there may not be logical grounds for assuming such a compensatory relationship between positive and negative indicators.

Representativeness is the heuristic in which the probability of an event is judged by the similarity of the event to the relevant population of events without regard to other factors. One effect of representativeness is that base-rate information may be ignored in diagnosis, although in prescriptive norms, such as Bayes' theorem, it is included in the calculation. Here, we also aim to study physicians' use/disregard of base-rate information in medical decision-making.[1]

[1] We wish to thank David O'Brien, Paul Pollard, and an anonymous reviewer of the grant supporting this research for noting a relationship between augmentation and representativeness as reflected in use of base-rate information. That is, in one case (representativeness) ignoring base rate is treated as bias, whereas in the other case (augmentation), taking base rate into account is treated as bias. Whether or not the contradiction is apparent or real is a controversial question beyond the scope of this paper to resolve. Indeed, in a recent critique Levi (1983) has argued that ignoring base rate is not an error according to Bayes' theorem! Without trying to resolve this question, we have attempted to show when and why base-rate information might be used differentially in the representativeness and augmentation problems.

Another effect of representativeness may be *redundancy*. In this case, the reasoner judges the probability of the event (disease) from the volume of diagnostic information *regardless of how redundant or non-independent the information is*. Reliance on redundant information is also related to another heuristic, *availability*, in which more graphically described symptoms may make the diagnosed disease condition more available in memory and imagination, and therefore judged more probable. In a slightly extrapolated sense, Eddy's study showed that physicians overprescribed mammograms because they felt more confident when they had reports of a mammogram, even when this information was redundant with other more definitive information available from a biopsy. The potential importance of this phenomenon becomes apparent when we consider the range of verbal fluency, boldness, etc. of patients or their representatives (e.g. parents or guardians of children).

We consider two forms of redundancy. In one the data are *literally* redundant within a similar (listed) format. In the other the two forms of the data do not differ in literal redundancy, but, rather, one is presented in a *contextualized* (paragraph) format, whereas the other is not. In this case the question concerns whether the contextualized form impresses the diagnostician as having more information.

Confusion of prospective-retrospective probabilities is the equating of the probability of the symptom given the disease with the probability of the disease given the symptom. It is worth noting that the former is the typical presentation in textbooks, lectures, and other didactic formats, and the latter is the typical presentation in a clinical setting.

Given these illustrations of the possible applicability of heuristics to medical diagnosis, two questions present themselves: (1) Is frequency of physicians' heuristic-based decisions specific to medical problems or is it a reflection of the widespread occurrence of heuristic-based biases so well-documented by Tversky, Kahneman, and others? (2) Does medical training affect the use of heuristics, and if so, does it impact differentially on medical–non-medical decision-making?

Method

Subjects. Forty-eight first-, second-, and third-year paediatric residents were recruited from two metropolitan hospitals. One hospital (A) is a medium-sized voluntary hospital in the Bronx, serving a poor and largely African–American and Hispanic population; the other hospital (B) is a somewhat larger voluntary hospital in Manhattan, serving a broader-spectrum population. Those residents (26 of 29 at Hospital A and 22 of 27 at Hospital B) who were available and agreed to participate were given questionnaires, comprised of pairs of hypothetical vignettes. Two of the 26 residents at Hospital A did not complete the questionnaires, and their data were discarded. One resident in Hospital A and one in Hospital B failed to indicate his year of residence, and their data were omitted from the analysis. This left a sample of 16 first-year residents (9 in hospital A, and 7 in hospital B), 14 second-year residents (7 in each hospital), and 14 third-year residents (7 in each hospital), for a total of 44 residents whose data were analysed.

Research Instruments. Each pair of vignettes was designed to assess a particular heuristic-based bias (Appendix A). One vignette within each pair tested for the use of the heuristic within a medical context, and the other within a non-medical context. The medical vignette within each pair duplicated (as much as possible) the structure of the non-medical vignette. The medical vignettes were taken from three modes of communication by which physicians acquire and transmit information about patients: (1) formal presentation during rounds, (2) communications about patients, during and after patient intakes, and (3) informal discussions about cases. The vignettes duplicated the details and order of data found in each of these modes as much as possible. The non-medical vignettes were mainly taken or adapted from the heuristics literature (e.g. Chapman & Chapman, 1969; Kahneman & Tversky, 1973; Oskamp, 1965; Tversky & Kahneman, 1982).

The instrument included nine pairs of hypothetical vignettes designed to test for the use of: (1) the discounting principle, (2) the augmentation principle, and (3) representativeness. In addition, physicians were tested for evidence of (4) confusion between prospective and retrospective probability. The medical and non-medical items were ordered randomly throughout the instrument, with the constraint that no two examples of the same heuristic appeared consecutively. One random order was used for all subjects.

In order to test for discounting, a strategy that terminates the search for alternative causes of an event once a plausible, sufficient cause has been identified, discounting items involved a list of symptoms of illness (medical context) and a list of automative dysfunctions (non-medical context), along with three diagnoses of each, with their characteristic symptoms. The subjects were instructed to assign probabilities to each of the diagnoses or combinations of diagnoses. Within each domain the three lists overlapped in part. On the medical problem the first and third (A + C) diagnoses *together* accounted for all the symptoms, and the second diagnosis (B) *by itself* accounted for all the symptoms. On the non-medical problem the second and third syndromes *together* and the first syndrome by itself accounted for the mechanical problem. Discounting was indicated when subjects failed to assign a probability greater than zero to the first and third diagnoses together.

In order to test for augmentation, a strategy based on the assumption that the occurrence of an event in the presence of a strong impeding factor necessarily implies the presence of a strong facilitating cause, the augmentation items required the respondents to indicate how much evidence would be necessary (i.e. weight loss [medical] and number of eyewitnesses [non-medical]) at three specified levels of confidence (10, 50, 90%) to diagnose anorexia nervosa (the medical problem) and commission of a violent crime (the non-medical problem). There were two versions of each problem. In one version of the case the patient/suspect was female (statistically more probable for a diagnosis of anorexia nervosa), and in the other the patient/suspect was male (more probable for violent crime). The relative base rates for anorexia nervosa and violent crime for males and females were stated explicitly. Augmentation was indicated by respondents' requiring a more stringent criterion for the presenting symptom (weight loss or eyewitnesses, respectively) when the base rate was lower than when it was higher.

Representativeness, a judgement strategy in which the probability of an event is judged by the similarity of the event to the relevant population of events *without regard to other factors*, was measured (1) by subjects' use or non-use of base-rate probability statistics, and (2) by subjects' overconfidence in redundant data as more informative. Vignettes testing the disregarding base-rate items required the subject to select a confidence level (0 through 100% in 11 equal-interval steps) in a diagnosis of a medical condition (pregnancy) or a non-medical condition (occupation). The medical and non-medical items were presented in two versions, one with a high and the other with a low explicitly stated base rate (30 vs. 70%). Disregard of base rate was indicated by an equal (or lower) probability estimate of the event in the higher compared to the lower base-rate version.

The prospective-retrospective items, testing for the equating of the probability of the symptom given the disease with the probability of the disease given the symptom, presented two versions each of a medical and a non-medical diagnostic problem: Kawasaki's Disease for the medical and salaries for physicians for the non-medical problems. One version required a prospective probability—e.g. the probability of the condition (Kawasaki's disease or being a physician) given the symptoms (3D rash, congestion, and a low-grade fever or making an annual salary of $40,000), and the other required a retrospective probability—i.e. the probability of the symptom given the condition. Confusion between prospective and retrospective probabilities was evident when subjects assigned exactly the same levels of probability for the prospective and retrospective events, which pretesting had shown to be substantially different.

The hypothesis that the paediatricians would have greater confidence in diagnoses based on redundant information was tested by comparing confidence levels assigned to each of three forms of a case study. (1) In the *non-redundant* listed form, 7 symptoms or behaviours were presented in a simple list. (2) In the *redundant listed* form, 12 additional symptoms or behaviours, which had been established as synonyms to the original set with a separate sample, were presented together with the original 7 to make a list of 19 indicators. (3) In the *contextually presented redundant* form the same 19 symptoms (in #2) were embedded in sentences within a coherent paragraph.[2]

The subject's task was to state a diagnosis, indicate what additional information would be needed, and indicate his/her level of confidence in the diagnosis by selecting one of ten alternatives (0–10%, 10–20%, . . . , 90–100%). Greater confidence in a diagnosis based on *redundant listed* data than in a diagnosis based on *non-redundant listed* data was expected. Greater confidence in decisions based on *contextualized redundant data* was indicated by a higher level of confidence in a diagnosis based on the contextualized redundant form than on the redundant listed form. This latter would indicate that contextual presentation over and above redundancy of the content also leads to increased confidence.

Procedure. Individual questionnaires were administered to all residents attending weekly staff conferences or Grand Rounds presentations. Residents who had not been present at staff meetings were given questionnaires individually.

Results

Data were inspected for similarity of distributions across the two hospitals. The two sets of data were similar for every measure despite the differences in the nature of the hospital, its staff, and the patient population. Therefore, we combined data from the two hospitals in further analyses. The data were analysed by means of a MANOVA to test for the effects of the measurement variation (to detect the basic phenomenon), by medical and non-medical contexts, level of training (first-, second-, and third-year residents), and the interaction among the variables. Significant interaction effects were followed up by Tukey tests of mean differences. A two-tailed test was used throughout.

---

[2] The synonyms were selected in a two-stage process: (1) *The Merriam–Webster Dictionary* (Wolf, 1974) and *Roget's Thesaurus* (Mawson, 1977) (for the non-medical symptoms), and the *New American Pocket Medical Dictionary* (Roper, 1978) (for the medical symptoms) were consulted for exact synonyms. However, the argument can be made that a pair of terms that may be synonyms in standard dictionaries may, in fact, convey different information to a particular sub-culture (e.g. a group of medical residents). Therefore a separate group of 10 medical residents was presented with the larger list of symptoms, including a large number of synonyms, and asked to circle those that "meant the same thing". Only those symptoms that all these pilot subjects agreed were equivalent were included on the final redundant list. Seven of the original nineteen terms were discarded as a result of this procedure.

Discounting. These data were analysed by the diagnosis chosen (2-item, 3-item, 5-item, and any combinations of these) by type of problem and year of residence. The interaction between diagnosis and type of problem was highly significant, $F(3, 123) = 767.33$, $p < 0.0001$. Decomposing this interaction revealed significant medical vs. non-medical differences in confidence in the 5-item, $F(1, 41) = 14.92$, $p < 0.0001$, 3-item, $F(1, 41) = 6.28$, $p < 0.02$, and 2-item diagnoses, $F(1, 41) = 9.10$, $p < 0.004$. These differences indicate that although there was a strong overall preference for the 5-item diagnosis, this preference was greater for the medical than the non-medical problem.

The medical–non-medical contrast was not significant for the probability of a multiple diagnosis (e.g. the 3-item + 2-item diagnosis). However, when we examine the data in terms of the frequency of discounting responses—i.e. of ignoring the possibility of more than one diagnosis being possible—70% (62/88) of the responses showed this pattern: 37/44 on the medical problem versus 25/44 on the non-medical problem. This difference is significant by z-test, $z = 1.98$, $p < 0.05$.

Augmentation. Overall, there was a large general effect for the expected and unexpected (gender), $F(1, 41) = 3011.76$, $p < 0.0001$, indicating the assumption of a compensatory relationship between a counter-indicator (unexpected gender) and the stringency of other diagnostic criteria (weight loss to diagnose anorexia nervosa for the medical problem, and number of witnesses to convict for a violent crime for the non-medical problem). There were no differences in the magnitude of the effect between the two kinds of problems (medical vs. non-medical) or year of residence, or the two in interaction.

Disregard of Base-Rate Information. Mean probability estimates are presented by type of problem and particular base-rate information provided in Table 1. There was a strong effect of introducing base-rate information, $F(4, 82) = 40.80$, $p < 0.0001$, but the effect was considerably less on the medical than the non-medical problem, $F(1, 35) = 25.03$, $p < 0.0001$, and this was superseded by an interaction effect between type of problem and year of residence, $F(2, 35) = 10.27$, $p < 0.001$. When the 30% vs. 70% contrast was made at each level, the following comparisons were significant: At all years on the non-medical problem and at year one on the medical problem, all p's < 0.001. The base-rate information was ignored, however, on the medical problem by second- and third-year residents. In particular, it was the lower base rate figure that was ignored or given little weight by the more advanced residents on the medical. On the non-medical problem, the probability estimates closely approximated the high and low base-rate figures.

Turning to the frequencies of this kind of heuristic-based bias by individual subjects, base-rate information was disregarded on the medical

TABLE 1
Probability Judgements According to Base Rate, Kind of Problem, and Year of Residence

	Problem					
	Medical			*Non-medical*		
	Base Rate			*Base Rate*		
Year of Residence	*30%*	*70%*	*Diff.*	*30%*	*70%*	*Diff.*
1	31.5	70.0	38.5	30.0	63.1	33.1
2	54.2	71.7	17.5	28.3	77.5	49.2
3	73.8	83.1	9.3	25.4	78.5	53.1
Mean	53.2	75.0	21.8	27.9	72.9	45.0

Note: Judgements are based on choice of one of 11 probabilities ranging from 0 to 100%.

problems by more of the third-year (10/14) than by second-year (5/12) or first-year residents (1/14). Further, disregard of base-rate information occurred in a far greater number of third-year residents' responses to the medical than to the non-medical problems (0/14), but this difference between domains did not occur for first- and second-year residents.

Redundancy. Mean confidence judgements are presented according to format of the information presented (non-redundant, listed redundant, contextualized redundant), type of problem, and year of residence in Table 2. There was a large difference between confidence in the non-redundant and the listed redundant information, $F(1, 41) = 428.26$, $p < 0.0001$, and a smaller but still significant difference between the redundant information when presented in listed and contextualized format, $F(1, 41) = 9.83$, $p < 0.003$. Further, there were highly significant interactions between the redundancy factor and medical–non-medical content, $F(2, 82) = 262.65$, $p < 0.0001$, and among the redundancy of the information, content of the problem, and year of residence, $F(4, 82) = 10.11$, $p < 0.0001$. The non-redundant vs. listed redundant, and the listed redundant vs. contextualized redundant were next separately analysed for effects of type of problem and year of residence. Neither type of problem nor years of training effects were obtained for redundant vs. listed redundant comparison, but they were significant for the listed vs. contextualized redundant comparison. Superseding these was a three-way interaction among the (listed vs. contextualized) format, type of problem, and year of residence, $F(2, 41) = 5.82$, $p < 0.006$. This was entirely due to greater confidence on the redundant contextualized version of the

TABLE 2
Probability Judgements By Format, Type of Problem, and Year of Residence

	Problems					
	Medical			*Non-medical*		
	Format			*Format*		
Year of Residence	*Non-redundant*	*Redundant*	*Contex-tualized*	*Non-redundant*	*Redundant*	*Contex-tualized*
1	12.5	38.8	45.0	12.5	39.4	40.0
2	10.0	31.4	32.1	13.6	36.4	38.6
3	10.0	27.1	27.1	12.9	38.6	37.1
Mean	10.9	32.7	35.2	13.0	38.2	38.6

Note: Judgements are based on choice of one of 10 probability intervals ranging from 0 to 10% through 90 to 100%, with the lower limit of the interval counted as the score.

medical problem by first-year than the two other groups of residents, $F(1, 15) = 12.10$, $p < 0.003$.

Confusion of Prospective and Retrospective Probability. For confusion of prospective and retrospective probability, a simple count was made of the number of instances in which the two probabilities were equal given the most relevant data since (1) there was no way to establish any alternative criterion for establishing how much of a difference would constitute confusion between the two, and (2) to equate the two problems as to the actual inequality of the pro- and retrospective probabilities.

Overall, 36% of the responses (32/88) showed evidence of this confusion, but there were no differences by level of training or problem context. Thus the probability of Kawasaki's disease/being a physician given a set of symptoms/making $40,000 annually was judged to be exactly the same as the reverse of these probabilities by over a third of the physicians, despite the obvious difference between the retrospective and prospective probabilities.

Discussion

The questions motivating this study were whether: (1) physicians show evidence of heuristic-based biases in solving problems concerning medical diagnoses; (2) the frequency or magnitude of these heuristic-based biases depends on the kind of diagnostic problem, medical or non-medical, and (3) the frequency or magnitude of biases is related to the diagnostician's level of hospital training. The answer to all three questions is a qualified "yes". All types of heuristic-based biases under study occurred, though the rate varied across heuristics and in relation to the content of the problem (medical–non-medical) and level of training.

Of particular interest were the findings that three types of heuristic-based biases were more evident on medical than non-medical problems, where we would assume the respondents had greater experience and specific training. Two of these effects occurred in interactions with level of training. The first medical–non-medical difference involved overconfidence in diagnosis depending on the format with which the information was presented. First-year residents were more confident in a diagnosis when the medical data were presented in context (embedded in paragraph) than in listed form, whereas second- and third-year residents were not. Training level did not affect occurrence on the non-medical problem. As there was no such interaction effect for the non-redundant–listed and redundant–listed conditions, the former finding was not due to redundancy alone, but to the embedding of the information in a paragraph or in a list. A simple explanation for this may lie in the nature of charts, where information is typically presented in listed format, and with the possibility that between the first and second year of residence physicians-in-training may become more familiar and comfortable with them.

It is also worth noting that the overall effect of redundancy itself, i.e. as indicated by the non-redundant–listed and redundant–listed presentations, was substantial (over 98%). Oskamp (1962) also found greater confidence with increasing amounts of data, even when the additional data were redundant. The practical implications in the medical domain are far-reaching. This implies that better-educated, more verbally fluent, more expressive patients may induce diagnoses that are more confidently held by the medical diagnostician than less-educated, less verbally fluent, and more diffident patients.

The second medical–non-medical difference occurred in discounting and indicated more use of this heuristic by first- than by second- or third-year residents, though only on the medical problems, and the difference among training levels was not very large. How is the effect of medical–non-medical content of the problem to be explained? The finding may become clearer if discounting is re-examined in terms of its underlying diagnostic hypotheses, as suggested by Braine, Connell, Freitag, and O'Brien (1990). Let p = the diagnosis and q = the presenting symptoms. Within this framework, discounting indicates the realization that if p then q does not imply that q if and only if p (i.e. p is a sufficient *and* necessary cause of q), but allows for the possibility that there may be other possible causes of q. Thus, the proper conclusion from if p then q is that one cannot draw a definitive and unique conclusion as to causation, i.e. indeterminacy.

In a study of evaluations of mechanical and medical causal and conditional hypotheses with undergraduate students as subjects, O'Brien, Costa, and Overton (1986) showed that the correct conclusions of indeterminacy as to causation given p and q was greater on medical than mechanical problems. The content differences were taken as indicating certain assumptions made about the nature of the mechanical (physical) and the medical (biological) world with greater indeterminacy ascribed to the biological than to the physical world.

The fact that discounting occurred more frequently on the medical problem than non-medical problems (which also was a mechanical problem in the present study) may thus be equivalent to the conclusion that indeterminacy was less often ascribed on the medical than the non-medical problem. On the face of it, the findings by O'Brien and his co-workers and ours are contradictory. However, we must remember that in their study undergraduate students were the problem-solvers whereas our subjects were advanced medical students (hospital residents). Thus, the studies are not strictly comparable. Nonetheless, their report does suggest considering the variations in frequency of discounting across types of problems and levels of expertise in terms of the underlying logical reasoning, which itself may be influenced by beliefs about the nature of the worlds (physical, biological, social, psychological . . .).

The third medical–non-medical effect involved the disregarding of base-rate information, which we assumed reflected representativeness. This finding shows that those presumably with greater expertise (third-year residents) disregarded base-rate information more than those with less expertise (first-year residents) but only in judging the medical problems, not the non-medical ones. This finding contrasts with the results of Tversky, Kahneman, and their collaborators (e.g. Kahneman, Slovic, & Tversky, 1982), who compared heuristic-based errors made by those with and those without special training in statistics. Their findings indicated that differences between these two groups were infrequent and relatively small. Where differences did exist, it was the experts who made fewer heuristic-based errors (see, however, Carroll, 1980). In contrast to our findings, Zukier and Pepitone (1984) found a greater utilization, in fact an overutilization, of base rate information by medical residents than by medical students. However, it should be noted that in their study Zukier and Pepitone used non-medical rather than medical problems to detect use of base rate information.

Before proceeding on to possible theoretical explanations of this interaction effect, we should consider the possibility that slight differences in the format of the medical and non-medical vignettes might explain the results. The base-rate information was presented within the main descriptive paragraph of the non-medical problem but separately and after the descriptive paragraph of the medical problem. Could this format difference account for the results?

We think it is highly unlikely, for the following reason: One would think that presenting the base-rate information separately and just prior to the request for a probability judgement in the medical problem rather than embedded in the paragraph as in the non-medical problem would tend to *highlight* the relevance of the base-rate information in the medical version of the problem, and yet it was precisely here that the base-rate information was overlooked (by third-year residents).

A more substantive explanation of the base-rate finding rests on a consideration of the intersection among prototypes, representativeness, and the character of medical training, as usually practised. Representativeness may rest, in part, on an overreliance on prototype matching (e.g. Cantor & Mischel, 1977; Rosch & Mervis, 1975; Sherman & Corty, 1984). From this perspective, categorization and recognition may involve constructing a typical case (or prototype) from one's experience and matching the current input against this prototype. Greater reliance on prototype matching should promote greater reliance on representativeness. Much of residents' medical training consists of exposure to, even drilling in, prototypes of particular disease conditions and syndromes as represented in medical textbooks and case studies. Such training may increase the strength of prototypes and result in an increase in the effects of representativeness. Of course, with further training and experience, involving exposure to greater variability and "the exceptions that make the rules", the constructed prototype may be more abstract so as to accommodate more variability.

An alternative explanation is that these training differences are not because of increasingly strong (medical) prototypes, but rather because the physician increasingly relies on prototype-matching (and therefore representativeness) with training. These alternative explanations could be tested by (1) assessing physicians' prototypes of different disease condition—some within and some outside their speciality, and (2) prototypes of specially constructed bogus diseases. If the former explanation is correct, then we would expect greater disregard of base rate only when the physicians are dealing with problems with which they are familiar (i.e. problems within their own speciality). If the latter explanation is correct, then more experienced diagnosticians would also show disregard of base rate (representativeness) on problems outside their own speciality and even when dealing with specially constructed (bogus) medical problems.

Another apparent contradiction in the findings is between the results for augmentation and disregarding base rate (see Footnote 1). On the one hand, we have asserted that the more experienced residents disregard base-rate information more than the less experienced on the medical (but not the non-medical) problem. On the other hand, we have asserted that there is an overwhelming tendency for our subjects (75%) to use augmentation whereby subjects *do use* base-rate information and (inappropriately) compensate for it in setting diagnostic criteria. A possible explanation is that problem-solvers have been observed to use base-rate information when the base rates are extreme enough and causally related (vs. arbitrary) (Ajzen, 1977). The extremities of the base rates were great in both sets of problems (augmentation and disregarding base rate), but they were more extreme in the former than in the latter. Further, whereas the base-rate difference is arbitrary in the situation designed to detect disregarding base rate—e.g. *at that particular conference* the proportion of accountants/salesmen was unequal—the base-rate difference on the augmentation problem was reflective of reality and possibly viewed as causal or at least not arbitrary.

Finally, although we are encouraged by the similarity of the results for the two hospitals from which the subjects were recruited, despite differences in their location, staff, and patient populations, it remains to be seen whether the findings will generalize across a broad spectrum of medical specialities and levels of practice. For example, it would be worthwhile to extend the range of training/expertise to post-residence. It is possible that the greater bias shown by the more experienced physicians occurs only within a small range of training. For example, attending physicians with even more experience may construct prototypes that accommodate a wider range of actual cases, thus limiting the logical error in disregarding base rate.

Sensitivity to patient privacy and the need for standardized situations led us to rely on a pencil-and-paper instrument. It could be argued that differences do exist between this format and "real-life decision-making"; we offer the following points: (1) Physicians are often asked to make decisions in response to patient-absent information (e.g. second opinions, test-taking situations) and (2) at least one study (Kirwan et al., 1983) has found similarity between medical reasoning on hypothetical and actual cases.

Our general conclusion is that heuristic-based biases often occur in diagnostic decision-making and their occurrence varies across non-medical versus medical contexts and according to level of training, in some instances increasing with medical training. This is open to at least two interpretations: (1) It may be explained by increasingly strengthened prototypes. (2) Training may increase reliance on heuristics such as representativeness that lead to increasing error.

REFERENCES

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology*, *35*, 303–314.

Braine, M.D.S., Connell, J., Freitag, J., & O'Brien, D.P. (1990). Is the base rate fallacy an instance of asserting the consequent? In K. Gilhooly, M. Keane, R. Logie, & G. Erdes (Eds.)

Lines of thought: Reflections on the psychology of thought. Vol. 1 (pp. 165–180). Chichester, England: Wiley.

Cantor, N. & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology, 35*, 38–48.

Cantor, N., Smith, E.E., French, R. des., & Mezzich, J. (1980). Psychiatric diagnosis as prototype categorization. *Journal of Abnormal Psychology, 89*, 181–193.

Carroll, J.S. (1980). Analyzing decision behavior: The magician's audience. In T.S. Wallgren (Ed.), *Cognitive processes in choice and decision behavior* (pp. 69–76). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Chapman, L.J. & Chapman, J.P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 72*, 193–204.

Chi, M., Glaser, R., & Farr, M.J. (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Eddy, D.M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgments under uncertainty: Heuristics and biases* (pp. 249–267). New York, Cambridge University Press.

Johnson, P.E., Duran, A.S., Hassebrock, F., Moller, J., Prietula, M., Feltovich, P.J., & Swanson, D.B. (1981). Expertise and error in diagnostic reasoning. *Cognitive Science, 5*, 235–283.

Joseph, G.-M. & Patel, V.L. (1990). Domain knowledge and hypothesis generation in diagnostic reasoning. *Medical Decision Making, 10*, 31–46.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 3*, 237–251.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kelley, H.H. (1972). Causal schemata and the attribution process. In E.E. Jones, D.E. Kanouse, H.H. Kelley, R.E. Nisbett, S. Valins, & B. Weiner. (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press.

Kruglanski, A.W., Friedland, N., & Farkah, E. (1984). Lay persons' sensitivity to statistical information: The case of high perceived applicability. *Journal of Personality and Social Psychology, 46*, 503–518.

Kirwan, J.R., Chaput de Saintonge, D.M., Joyce, C.R.B., et al. (1983). Clinical judgment in rheumatoid arthritis. I. Rheumatologists' opinions and the development of "paper patients". *Annual of Rheumatoid Disorders, 42*, 648–651.

Levi, I. (1983). Who commits the base rate fallacy? *The Behavioral and Brain Sciences, 3*, 502–506.

Mawson, C.O.S. (Ed.) (1977). *Roget's Thesaurus*. New York: Simon & Shuster.

Murphy, G.L. & Wright, J.C. (1984). Changes in conceptual structure with expertise: Differences between real-world experts and novices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 144–155.

O'Brien, D.P., Costa, G., & Overton, W.F. (1986). Evaluations of causal and conditional hypotheses. *Quarterly Journal of Experimental Psychology, 38*, 493–512.

Oskamp, S. (1962). The relationship of clinical experience and training methods to several criteria of clinical prediction. *Psychological Monographs, 76* (28, whole No. 547).

Patel, V.L. & Groen, G.J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science, 10*, 91–116.

Rosch, E. & Mervis, C.B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573–605.

Roper, N. (Ed.) (1978). *New American Pocket Medical Dictionary*. New York: Churchill Livingstone.

Ross, L. (1981). The "intuitive scientist" formulation and its developmental implications. In J.H. Flavell & L. Ross (Eds.), *Social cognitive development: Frontiers and possible futures* (pp. 1–42). Cambridge: Cambridge Universities Press.
Sherman, S.J. & Corty, E. (1984). In R.S. Wyer, Jr. & T.K. Srull (Eds.), *Handbook of social cognition, Vol. 1* (pp. 189–286). Lawrence Erlbaum Associates, Inc.
Tversky, A. & Kahneman, D. (1982). Judgment of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgments under uncertainty: Heuristics and biases* (pp. 84–98). New York: Cambridge University Press.
Wolf, H.B. (Ed.) (1974). *The Merriam-Webster Dictionary.* New York: Simon & Shuster.
Zukier, H. & Pepitone, A. (1984). Social roles and strategies in prediction: Some determinants of the use of base-rate information. *Journal of Personality and Social Psychology, 47,* 349–360.

Manuscript received 7 January 1991

APPENDIX A

Medical and Nonmedical Vignettes

Discounting Vignette (Non-medical)

You have been driving your car on the highway for about an hour and a half. The temperature gauge indicates that the engine is overheated. You see steam coming from under the hood. You smell antifreeze. You get out of the car and notice that the radiator fluid is overflowing. Two fresh puddles are seen under the car: one under the radiator and one under the oil pan.

Three car problems called, 1, 2, and 3, are listed with their attendent signs.

Problem 1	*Problem 2*	*Problem 3*
Temp. gauge indicates engine is overheated	Steam coming from under hood	Smell of antifreeze
Radiator fluid overflow	Two fresh puddles; one under radiator, one under oil pan	Radiator fluid overflow
Smell of antifreeze		Temp. gauge indicates engine is overheated
Steam coming from under hood		
Two fresh puddles; one under radiator, one under oil pan		

Discounting Vignette (Medical)

A 7-month-old is brought in by the mother who complains of the following symptoms: wheezing, a "cold" for two or three days, difficulty in breathing. He responds minimally to plain mist. He responds well to 0.1% epinephrine.

Three clinical syndromes called A, B, and C, are listed with their attendent symptoms and response to treatment.

Syndrome A	*Syndrome B*	*Syndrome C*
Wheezing	Wheezing	Cold (2–3 days)
Difficulty breathing	Cold (2–3 days)	Responds to 0.1% epin.
Min. resp. to plain mist	Responds to 0.1% epin.	
	Difficulty breathing	
	Min. resp. to plain mist	

For both versions, subjects were asked to:

1. List the probability that the above patient has syndrome A, B, or C, or any combination of syndromes (based on a total of 100%).

Please feel free to explain your answer or offer any additional information.

____

____

____

Augmentation Vignettes (Non-medical)

Males constitute 97% of those convicted of violent crimes.

You are a juror judging the guilt or innocence of a *male*. He is accused of a violent crime. How many witnesses' testimony would you require to be convinced of his guilt?

1. How many witness(es)' testimony would you require to be convinced of his guilt (your confidence level = 10%)?_____

2. How many witness(es)' testimony would you require to be convinced of his guilt (your confidence level = 50%)?_____

3. How many witness(es)' testimony would you require to be convinced of his guilt (your confidence level = 90%)?_____

Please feel free to explain your answer or offer any additional information.

____

____

____

For comparison, and in order to determine augmentation, the above vignette was repeated and subjects were presented with the same percentage of convictions for males. In this version of the vignette, however, subjects were asked to respond to all questions listed above while judging the guilt or innocence of a female.

Augmentation Vignettes (Medical)

Females comprise about 97% of patients diagnosed with Anorexia nervosa.

A 14-year-old *female* is seen:

1. What percentage of weight loss would she need to exhibit for a diagnosis of Anorexia nervosa (your confidence level = 10%)?_____
2. What percentage of weight loss would she need to exhibit for a diagnosis of Anorexia nervosa (your confidence level = 50%)?_____
3. What percentage of weight loss would she need to exhibit for a diagnosis of Anorexia nervosa (your confidence level = 90%)?_____

Please feel free to explain your answer or offer any additional information.

____

____

____

____

For comparison, and in order to determine augmentation, the above vignette was repeated, and subjects were presented with the same percentage rates of this disorder seen in females. In this version of the vignette, however, subjects were asked to respond to all questions above while diagnosing a male.

Disregard of Base-Rate Information (Non-medical)

Version 1:

You are attending a conference, and you enter a reception area in which you know that 70% of the people are salesmen and the remainder are accountants. A man approaches you and appears quite friendly. He attempts to involve you in conversation. You answer his question tersely, but he appears to not notice. The only way you can stop the conversation is to leave the room.

1. What is the probability that this man is a salesman?

   0% 10% 20% 30% 40% 50%
   60% 70% 80% 90% 100%

Please feel free to explain your answer or offer any additional information.

Version 2:

For comparison, and in order to determine disregarding of base-rate, the same vignette was repeated with one difference: the percentage of salesmen was equal to 30% (as opposed to 70% above). The same questions were presented.

Disregard of Base-Rate Information (Medical)

Version 1:

You are working in a walk-in paediatric clinic in which adolescents are seen to age 18.

A sixteen-year-old girl has been found to have an abnormal urinalysis. Bacteria and W.B.C.'s were seen as on a clean-catch urinalysis. She has had a positive U.C.G., but states that she had a normal menstrual period about a week ago. She complains of a bit of nausea in the morning and states that she feels like throwing up. She denies sexual activity.

The pregnancy rate among adolescents at this clinic is seventy percent (70%).

1. What is the probability that this patient is pregnant?
   0% 10% 20% 30% 40% 50%
   60% 70% 80% 90% 100%

Please feel free to explain your answer or offer any additional information.

Version 2:

For comparison, and in order to determine the disregarding of base-rate, the same vignette was repeated with one difference: the pregnancy rate among adolescents at this clinic is 30% (as opposed to 70% above). The same questions were presented.

Redundancy Vignettes (Non-medical)

Simple Version:

Tom Wilson is of high intelligence although lacking true creativity. He is neat, organized, and prefers orderly systems. He is generally introverted and unemotional.

Listed Version:

Stan Thompson is highly intelligent, learns quickly, is well able to repeat much of what he has read, does not talk about it a great deal, remembers much of what he is taught, is intelligent, ability to produce original thought appears limited, does not produce much original work, is relatively unimaginative, is neat, likes to keep things tidy, prefers orderly systems, arranges things around him in an organized manner, is quiet, not prone to shows of emotion, doesn't take part in much of the conversation around him and appears unemotional.

Contextual Version:

Using name Alan Green, this version restates the listed version in sentence form.

For all versions, subjects were asked to:

1. Circle the graduate program that Tom (Stan, Alan) would be most likely to enroll in:

   Business Administration
   Computer Science
   Engineering
   Medicine
   Humanities and Education
   Physical and Life Sciences
   Law
   Library Sciences
   Social Science and Social Work

2. What additional information, if any, would you like to have?

   ____

   ____

3. What is your level of confidence?
   (If more than one choice, indicate level of confidence for each)

   0–10% 10–20% 20–30% 30–40% 40–50%
   50–60% 60–70% 70–80% 80–90% 90–100%

Please feel free to explain your answer or offer any additional information.

Redundancy Vignettes (Medical)

Simple Version:

A ten and a half-year-old female presents with loss of appetite, fatigue, abdominal pain, nausea, vomiting, fever 100.6° F, and insomnia due to pain.

Listed Version:

An eleven-year-old female presents with nausea, vomiting, "regurgitation", "feeling like throwing up", "felt like she did when she was seasick", loss of appetite, "wasn't hungry", fever 100.4° F, "feels hot", fatigue, "tiredness", "lack of energy", insomnia, "not sleeping well", abdominal pain, and "pain in stomach", "a belly-ache".

Contextual Version:

Same as listed version in sentence form.

For all versions, subjects were asked:

1. What is your tentative diagnosis(es)?

2. What additional information, if any, would you need from history and/or physical and/or diagnostic tests?

3. What is your level of confidence at this time?
   (If more than one choice, indicate level of confidence for each)

   0–10% 10–20% 20–30% 30–40% 40–50%
   50–60% 60–70% 70–80% 80–90% 90–100%

Please feel free to explain your answer or offer any additional information.

Confusion of Prospective and Retrospective Probability (Non-medical)

Version 1:

What is the probability that a person who will earn $40,000 per year will have completed medical school?

0–10%	10–20%	20–30%	30–40%	40–50%
50–60%	60–70%	70–80%	80–90%	90–100%

Version 2:

What is the probability that a person who completes medical school will earn $40,000 per year?

0–10%	10–20%	20–30%	30–40%	40–50%
50–60%	60–70%	70–80%	80–90%	90–100%

In both versions, the following was added:

Please feel free to explain your answer or offer any additional information

____

____

____

____

Confusion of Prospective and Retrospective Probability (Medical)

Version 1:

A 3-year-old presents with a rash, red fissured lips and a fever.

1. What is the probability that this child has Kawaski's Disease?

0–10%	10–20%	20–30%	30–40%	40–50%
50–60%	60–70%	70–80%	80–90%	90–100%

Version 2:

A 3-year-old child is diagnosed as having Kawaski's Disease

1. What is the probability that this child presented with a rash, red fissured lips and a fever?

0–10%	10–20%	20–30%	30–40%	40–50%
50–60%	60–70%	70–80%	80–90%	90–100%

In both versions, the following was added:

Please feel free to explain your answer or offer any additional information.

____

____

____