ELSEVIER

COGNITION

Cognition 76 (2000) 179–193

www.elsevier.com/locate/cognit

Gigerenzer's normative critique of Kahneman and Tversky

Peter B.M. Vranas

Department of Philosophy, The University of Michigan, 2215 Angell Hall, Ann Arbor, MI 48109, USA

Received 16 October 1999; accepted 14 December 1999

Abstract

Gigerenzer has argued that it may be inappropriate to characterize some of the biases identified by Kahneman and Tversky as "errors" or "fallacies," for three reasons: (a) according to frequentists, no norms are appropriate for single-case judgments because single-case probabilities are meaningless; (b) even if single-case probabilities make sense, they need not be governed by statistical norms because such norms are "content-blind" and can conflict with conversational norms; (c) conflicting statistical norms exist. I try to clear up certain misunderstandings that may have hindered progress in this debate. Gigerenzer's main point turns out to be far less extreme than the position of "normative agnosticism" attributed to him by Kahneman and Tversky: Gigerenzer is not denying that norms appropriate for single-case judgments exist, but is rather complaining that the existence and the nature of such norms have been dogmatically assumed by the heuristics and biases literature. In response to this complaint I argue that single-case probabilities (a) make sense and (b) are governed by probabilistic norms, and that (c) the existence of conflicting statistical norms may be less widespread and less damaging than Gigerenzer thinks.

Keywords: Heuristics and biases; Judgment under uncertainty; Norms of reasoning; Frequentism; Concepts of probability

1. Introduction

1.1. Gigerenzer's three critiques

In a series of publications, Gigerenzer (1991a,b; 1993; 1994; 1996; 1998) and his colleagues (Gigerenzer, Hell & Blank, 1988; Gigerenzer & Hoffrage, 1995; Gigerenzer, Hoffrage & Kleinbölting, 1991; Gigerenzer & Murray, 1987, chap. 5; Giger-

E-mail address: vranas@umich.edu (P.B.M. Vranas).

0010-0277/00/$ - see front matter
PII: S0010-0277(99)00084-0

enzer, Swijtink, Porter, Daston, Beatty & Krüger, 1989, chap. 6; Gigerenzer, Todd & ABC Research Group, 1999; Hertwig & Gigerenzer, 1999; see also Cosmides & Tooby, 1996) have expressed some reservations about a program of research which was initiated by Kahneman and Tversky (Kahneman, Slovic & Tversky, 1982; Tversky & Kahneman, 1974; 1983; see also Nisbett & Ross, 1980; Piattelli-Palmarini, 1994) and which has come to be known as "the heuristics and biases approach to judgment under uncertainty". Gigerenzer's reservations can be divided into three groups. First, on the *empirical* level, Gigerenzer argues that some of the biases identified by Kahneman and Tversky are unstable, in the sense that for example in some cases their magnitude can be considerably reduced by asking questions in terms of frequencies rather than in terms of probabilities. Second, on the *methodological* level, Gigerenzer argues that, because Kahneman and Tversky's heuristics are formulated by means of vague, atheoretical terms like "representativeness," the appeal to these heuristics as generators of biases has limited explanatory power; Gigerenzer advocates instead an increasing emphasis on investigating the cognitive *processes* that underlie judgment under uncertainty. Third, on the *normative* level, Gigerenzer argues that it may be inappropriate to characterize some of the biases identified by Kahneman and Tversky as "errors" or "fallacies," for three reasons. (a) According to frequentists, no norms are appropriate for single-case judgments because single-case probabilities are meaningless. (b) Even if single-case probabilities make sense, they need not be governed by statistical norms because such norms are "content-blind" and can conflict with conversational norms. (c) In some cases conflicting statistical norms exist ("statistics does not speak with one voice").

Because it turns out that Gigerenzer's *empirical* disagreement with Kahneman and Tversky is probably much smaller than one might think at first sight (Kahneman & Tversky, 1996), and because I believe that Gigerenzer's *methodological* remarks are on the right track, I will focus on Gigerenzer's *normative* critique, which has been so far relatively neglected. I will try to clear up certain misunderstandings which may have hindered progress in the debate, and I will argue that some of Gigerenzer's reservations can be countered: it does seem appropriate to characterize some of the biases identified by Kahneman and Tversky as "errors" or "fallacies". In the next three sections I address the specific points of Gigerenzer's normative critique; in the remainder of the current section I explain why normative issues are worthy of investigation.

1.2. The importance of normative issues

Given that, as Kahneman and Tversky note, "Gigerenzer's [normative] critique consists of a conceptual argument against our use of the term "bias"," (Kahneman & Tversky, 1996, p. 582) some people may take the dispute to be merely terminological. Although at a superficial level the dispute is indeed terminological, at a deeper level what matters is Gigerenzer's *reason* for objecting to the use of the term "bias": Gigerenzer argues that Kahneman and Tversky may be comparing the performance of the participants in their experiments with *inappropriate (i.e. incorrect) norms.* But the question of which – if any – norms are appropriate for judgment under uncertainty is far from terminological and may be of the utmost importance: given that uncertainty is a pervasive fact of life, making the *right* judgments under conditions of uncertainty has practical significance. The topic of judgments under uncertainty is not exhausted by describing the relevant cognitive processes: one also wants to know how to *improve* such judgments, and improvement presupposes existence of appropriate norms. The issue of whether such norms exist is worthy of investigation, as is the issue of whether the norms delivered by probability and statistics are appropriate. Given the nature of these issues, the investigation will have to be primarily conceptual rather than empirical, but conceptual clarification is a prerequisite for empirical progress.[1]

2. Does frequentism cast doubt on the existence of appropriate norms?

2.1. Gigerenzer's argument

For present purposes, it suffices to define roughly a *single-case proposition* as a proposition about a particular individual or event (e.g. the proposition that Linda is a bank teller) and a *single-case judgment* as a judgment expressing one's (degree of) confidence in (the truth of) a single-case proposition (e.g. the judgment that one is 75% confident that Linda is a bank teller). Let P be:

(P) Probabilities cannot be meaningfully applied to single-case propositions.

Gigerenzer has repeatedly advanced an argument that can be formulated as follows (Gigerenzer, 1991a, p. 88; 1996, p. 593; Gigerenzer et al., 1991, p. 525):

(FP) According to frequentism, P is true.

Thus: (FC) According to frequentism, no norms are appropriate for single-case judgments.

Some of Gigerenzer's writings may give the impression that he also endorses the further argument from FC and P′ to C:

(P′) Frequentism is true.

Thus: (C) No norms are appropriate for single-case judgments.

Indeed, Gigerenzer (1991a, p. 88) claims that "the frequentist interpretation of probability has been dominant since about 1840" and "dominates today's statistics departments" (p. 87; cf. Cosmides & Tooby, 1996, p. 4; Gigerenzer et al., 1991, p. 525). In other places, however, Gigerenzer is careful to avoid assuming that frequentism is true (Gigerenzer, 1993, p. 290; 1994, p. 141) and stops short of asserting C, limiting himself instead to FC (Gigerenzer, 1991a, p. 92; Gigerenzer et al., 1991, p. 525). In fact, Gigerenzer (pers. commun.) has confirmed that he does not intend to assert P′ or to infer C from FC. This is an important point at which a lack of clarity in the debate has led the participants to talk at cross-purposes (e.g. Samuels, Stich & Bishop, in press): Kahneman and Tversky (1996, p. 586) attributed to Gigerenzer the position of "normative agnosticism" (namely C), but Gigerenzer

[1] The above remarks suggest an answer to the question of Gigerenzer and Murray (1987, p. 179): "Why have experimental psychologists become so involved with how people *should* think, even though prescription is atypical of experimental psychology?"

in his reply (Gigerenzer, 1996) did not explicitly disavow this position (he disavowed it in a personal communication, 1999). In any case, more important than who said what is the clarification of Gigerenzer's position: Gigerenzer is not denying the existence of norms that are appropriate for single-case judgments, but is rather arguing that the widespread acceptance of frequentism *places the burden of proof* on those (like Kahneman and Tversky) who claim that such norms exist.

I deny that the predominance of frequentism generates such a burden of proof: in response to Gigerenzer's argument, in the next two subsections I argue respectively that (a) properly understood, P is false, and that (b) FC does not follow from FP because C does not follow from P.

2.2. Frequentism and single-case probabilities

How could anyone hold P? According to the standardly accepted definition, a probability measure is any real-valued, nonnegative, countably additive function that has as domain a σ-field of subsets of a sample space and assigns value 1 to the sample space (e.g. Dudley, 1989, p. 196).[2] Provided that propositions can be modeled as sets (e.g. as sets of "possible worlds" (Lewis, 1986, p. 53)), nothing prevents one from defining probability measures on single-case propositions (or on sets of apples for that matter). No frequentist can reasonably deny this; so frequentists who assert P probably mean something else.

To see what they might mean, it is essential to distinguish the (above-mentioned) mathematical *definition* of a probability measure from various *concepts* of probability (or the probability *calculus* from various *interpretations* of that calculus; cf. Cosmides & Tooby, 1996, p. 8). A frequentist concept of probability is concerned with probability measures whose domains are sets of collections of events (or sets of multiple-case propositions) and whose values are (limits of) relative frequencies in these collections. A subjectivist concept of probability is concerned with probability measures whose domains are sets of propositions and whose values are degrees of confidence in these propositions. Further concepts of probability can be defined (Carnap, 1945; Good, 1959; Howson, 1995; Mackie, 1973, chap. 5). Now by definition frequentist probabilities cannot be meaningfully applied to single-case propositions but subjectivist probabilities can; therefore, frequentists who assert P by referring to "probabilities" simpliciter may be neglecting the fact that more than one concept of probability exists or may be assuming that non-frequentist concepts of probability are problematic.

But I see no obvious way in which a subjectivist concept of probability is problematic. Nothing prevents one from defining a function which to every proposition (including single-case ones) believed by an individual with a certain degree of confidence assigns a number in [0, 1] which is the individual's degree of confidence in the proposition. Nothing prevents this function from being a probability measure

[2] A σ-field is a set of subsets of the sample space that includes the sample space and is closed under complementation and countable union. A function f on a σ-field is countably additive exactly if, for any infinite sequence $\{A_n\}$ of mutually disjoint members of the σ-field, $f(\cup A_n) = \Sigma f(A_n)$.

in the mathematical sense, and thus a subjectivist concept of probability from being instantiated. I am not claiming that a subjectivist concept of probability *is* instantiated: maybe people's degrees of confidence never obey the probability axioms. I am claiming only that such a concept *can* be instantiated, and this is enough to refute P. Some frequentists might insist: "The phrase "probability of death", when it refers to a single person, has no meaning at all for us" (von Mises, 1928/1957, p. 11; quoted by Gigerenzer et al., 1991, p. 525; also by Gigerenzer, 1991a, p. 88; 1993, p. 292). If their claim is not about subjectivist probability (e.g. if they are denying that a particular person – or uranium atom – has a certain "objective probability" or "propensity" or "chance" of dying – or decaying – within the next year), then their claim is irrelevant to present concerns. But if their claim *is* about subjectivist probability, then the dispute becomes *merely* terminological. One could go along with frequentists and refuse to call degrees of confidence "probabilities";[3] the fact would remain that nothing prevents degrees of confidence from being probabilities in the mathematical sense and thus from instantiating a legitimate concept of probability.

In conclusion, if the "probabilities" referred to in P are frequentist, then P is true but irrelevant; and if they are subjectivist (i.e. degrees of confidence), then P is relevant but false.

2.3. Appropriate norms without subjectivist probabilities

Contrary to what we have been given reason to believe so far, suppose that subjectivist probabilities cannot be meaningfully applied to single-case propositions because a subjectivist concept of probability is problematic. Maybe, for instance, the sets of people's beliefs cannot correspond to σ-fields; or maybe people's beliefs cannot have precise numerical degrees of confidence. (I believe that these problems can be met, but I am making a counterfactual supposition for the sake of argument.) Still, it does not follow that no norms are appropriate for single-case judgments. One can still plausibly claim, for instance, that it is unreasonable to be highly confident both in a proposition and in its negation (cf. Kahneman & Tversky, 1996, p. 586). Of course the question arises how such norms are to be justified. I address this question in the next section; my present point is only that, since it is *possible* that norms appropriate for single-case judgments exist even if single-case probabilities make no sense, C does not deductively *follow* from P, so that FC does not follow from FP. In other words, the case for norms appropriate for single-case judgments does not stand and fall with the case for single-case probabilities. Actually Gigerenzer agrees: he understands "no norms" in C and FC as "no statistical norms" (pers. commun.). But Gigerenzer has not always been clear on this matter and has sometimes been misunderstood; for example, the position of "normative agnosticism" that Kahneman and Tversky (1996, p. 586) attributed to Gigerenzer – and that Gigerenzer (1996) in his

[3] Thus one might grant a frequentist request to replace, in psychological experiments, questions like "How probable is it that Linda is a bank teller?" with questions like "How confident are you that Linda is a bank teller?"

reply did not explicitly disavow – concerns the appropriateness for single-case judgments of *any* normative standards. In any case, the more important point is that the widespread acceptance of frequentism places no burden of proof on those who believe that norms appropriate for single-case judgments exist.

3. Arguments for the existence of appropriate norms

Even if Gigerenzer is wrong to claim that the predominance of frequentism casts doubt on the existence of norms appropriate for single-case judgments, arguably Gigerenzer is right to complain that the existence and the nature of such norms have been uncritically assumed by the heuristics and biases literature. In response to Gigerenzer's complaint, in the next two subsections I try to justify respectively (a) the *extra-statistical* norm of calibration and (b) the *probabilistic* norms (like the "conjunction rule") which specify that degrees of confidence ought to satisfy the probability axioms. (These probabilistic norms should be distinguished from *statistical* norms like the norm that, when all one knows about a ball is that it was randomly drawn from an urn containing two black and three white balls, one ought to be 40% confident that the ball is black. Statistical norms are dealt with in the next section.)

3.1. Is calibration an appropriate norm?

In calibration experiments, participants typically answer a series of questions and indicate, after answering each question, their degree of confidence in the correctness of their answer (Alpert & Raiffa, 1982; Keren, 1991; Lichtenstein, Fischhoff & Phillips, 1982). A typical finding, labeled *overconfidence*, is that confidence exceeds accuracy; i.e. for each degree of confidence x, the proportion of correct answers among the answers for which confidence x was indicated is lower than x. (E.g. among the answers for which confidence 90% was indicated, only 75% may be correct.) Gigerenzer (1991a, p. 88; 1993, p. 298) asks: "has probability theory been violated if one's *degree of belief (confidence) in a single event* (i.e. that a particular answer is correct) is different from the *relative frequency* of correct answers one generates in the long run?" Gigerenzer answers negatively, but his answer is irrelevant, because his *question* is misleading. In calibration experiments, the relative frequency of correct answers one generates in the long run is *not* compared with one's degree of confidence in a *single* event, but rather with a degree of confidence that happens to be the same for a *series* of single-case judgments:

> If a person assesses the probability of a proposition being true as 0.7 and later finds that the proposition is false, that in itself does not invalidate the assessment. However, if a judge assigns 0.7 to 10 000 independent propositions, only 25 of which subsequently are found to be true, there is something wrong with these assessments. The attribute that they lack is called calibration (Lichtenstein et al., 1982, p. 306; cf. Alpert & Raiffa, 1982, p. 295).

Gigerenzer might agree but still demand a justification for the norm of calibration. Here is an attempt to come by such a justification:

> A patient who is informed by his surgeon that she is 99% confident in his complete recovery may be justifiably upset to learn that when the surgeon expresses that level of confidence, she is actually correct only 75% of the time. Furthermore, we suggest that both surgeon and patient are likely to agree that such a calibration failure is undesirable (Kahneman & Tversky, 1996, p. 588).

Would Gigerenzer be indifferent between getting a diagnosis of almost certain recovery from a calibrated and from an overconfident surgeon? Would Gigerenzer be indifferent between getting a weather forecast of almost certain sunshine from a calibrated and from an overconfident forecaster? It seems that calibrated persons have both a pragmatic and an epistemic advantage. Gigerenzer (pers. commun.) has replied that "one cannot justify a general norm by providing a few reasonable examples". True, but the plausibility of these examples now shifts the burden of proof to those (like Gigerenzer) who *question* the norm of calibration. On the other hand, I grant that these examples are different from those with which participants in calibration experiments are presented, and thus do not straightforwardly support the claim that overconfidence in *experimental* contexts is an "error".

3.2. Arguments for the appropriateness of probabilistic norms

A simple consequence of the definition of a probability measure is that the probability of the conjunction of two propositions cannot exceed the probability of either conjunct. A corresponding norm for judgments under uncertainty is the *conjunction rule*: one's degree of confidence in the conjunction of two propositions ought not to exceed one's degree of confidence in (either) one of the conjuncts (e.g. Stein, 1996, p. 6). Tversky and Kahneman (1983) found that people's judgments frequently violate the conjunction rule. For example, when presented with a description suggesting that Linda is a feminist but not a bank teller, people typically assign higher confidence to the proposition that Linda is a bank teller and is active in the feminist movement than to the proposition that Linda is a bank teller. Gigerenzer (1991a, p. 91; 1996, p. 593) suggests that there is nothing wrong with such judgments. Gigerenzer, however, neglects to address the fact that two long-standing research programs have aimed at showing that degrees of confidence ought to satisfy the axioms of probability. The first of these programs starts with the assumption that the preferences of *rational* persons satisfy some simple conditions (like transitivity) and goes on to prove, by means of *representation theorems*, that rational persons have probability and utility functions with respect to which their preferences maxi-

[4] Gigerenzer might question the assumption that people's preferences ought to satisfy conditions like transitivity. But there is a voluminous literature on the justification of these conditions (e.g. McClennen, 1990), a literature that Gigerenzer fails to address. (Note that evidence that people's preferences *in fact* fail to satisfy these conditions does not in itself provide a ground for questioning the assumption that they *ought* to satisfy them.)

mize expected utility (Maher, 1993; Ramsey, 1926/1980; Savage, 1954).[4] The second of these programs tries to show, by means of *Dutch book arguments*, that people whose degrees of confidence violate the axioms of probability can be made to suffer a sure loss in a betting situation (de Finetti, 1937/1980; Horwich, 1982; Howson & Urbach, 1993; for a critique see Maher, 1993; 1997).[5] Recently a third research program has also gotten under way: Joyce (1998) has proven a theorem to the effect that, given certain reasonable constraints that any measure of accuracy for a system of beliefs must satisfy, for any set of degrees of confidence which violates the probability axioms there is another set which satisfies the axioms and is *more accurate* than the former set in *every* possible world (i.e. no matter which propositions turn out to be true and which false). So the epistemic goal of maximizing the accuracy of one's beliefs provides a justification of probabilistic norms. Gigerenzer adduces no reason for believing that any of these research programs fails; thus Gigerenzer is not entitled to conclude that there is nothing wrong with judgments which violate the conjunction rule, or, more generally, the probability axioms.

Responding to an earlier version of this paper, Gigerenzer (pers. commun.) claimed that I did "not provide a justification for why representation theorems and Dutch book arguments apply to the Linda problem". But by their very nature representation theorems and Dutch book arguments are *content-neutral*: they apply to *every* problem, and thus also to the Linda problem. Gigerenzer believes that "a norm needs to be justified for a given situation or problem" (pers. commun.). But then it is incumbent on him to explain what is wrong with the above content-neutral justifications of probabilistic norms; it is not enough to *assert* that norm justification cannot be content-neutral. Gigerenzer might respond that he does have an argument against the content-neutrality of norm justification, as follows. To solve textbook probability or statistics problems certain assumptions are standardly made which typically do not apply to real-world problems. For example, in real-world problems the available data are seldom the outcome of random sampling, and people are frequently able to search for information and "decide for themselves which features of the problem situation might be relevant" (Gigerenzer & Murray, 1987, p. 164; cf. Gigerenzer et al., 1989, p. 230). So the norms that apply to textbook problems differ from the norms that apply to real-world problems. This argument, however, does not show that in real-world problems one's degrees of confidence need not satisfy the probability axioms; the argument shows rather that, when applying statistical norms to a real-world problem, one needs to examine carefully the appropriateness of one's assumptions by taking into account the content of the problem. In *this* sense of "content-neutrality" I agree that the justification of *statistical* norms cannot be content-neutral; but this is a sense different from the one in which the justification

[5] Dutch book arguments claim that people whose degrees of confidence violate the probability axioms *can*, not that they *will*, be made to suffer a sure loss. One might thus object that these arguments give people little *pragmatic* reason to conform to the probability axioms. True, but the arguments may still give people *epistemic* reason: the fact that a smart enough opponent could defeat me suggests that my beliefs are less than ideal even if no actual opponent is smart enough. In any case, I am not claiming that Dutch book arguments are conclusive: I am claiming only that they are relevant and should thus be addressed by Gigerenzer.

of *probabilistic* norms provided by representation theorems and the like *is* content-neutral.

Gigerenzer has also replied that probabilistic norms need not apply to the Linda problem because conversational norms may have instructed the participants to understand "probability" non-mathematically (cf. Hertwig & Gigerenzer, 1999). I respond to an analogous point in the next section when I address the "Engineer/Lawyer problem".

4. Are statistical norms appropriate?

4.1. Real-world versus experimental contexts

As we saw at the end of the last section, Gigerenzer remarks that caution is needed when applying statistical norms to real-world contexts. Therefore, Gigerenzer and Murray argue, from the premise that participants in psychological experiments perform poorly the "conclusion that the man-in-the-street's reasoning is biased and irrational is … unwarranted" (Gigerenzer & Murray, 1987, p. 167).

Although *some* real-world contexts (e.g. gambling) closely approximate certain experimental contexts, it seems indeed that the inference from poor performance in experimental contexts to poor performance in real-world contexts is not immediate. (This is not to say that the inference is unwarranted; see Nisbett & Ross, 1980, chap. 11.) Moreover, as I said, I agree with Gigerenzer that one should not apply statistical norms in a mechanical fashion but should rather check whether the content of each particular real-world situation warrants such an application (for a concrete example, see the "cab problem" in the next subsection). On the other hand, note that performance in *experimental* contexts should be compared with norms applicable to *experimental* contexts in order to determine whether the biases identified by Kahneman and Tversky should be characterized as "errors" or "fallacies". Given that experimenters are typically careful to for example stipulate random sampling, Gigerenzer's points about real-world contexts provide in themselves no argument against the appropriateness of statistical norms for experimental contexts.[6]

Gigerenzer's writings might suggest, however, that statistical norms are sometimes inappropriate *even* for experimental contexts. Consider the Engineer/Lawyer (E/L) experiments (Kahneman & Tversky, 1973; cf. Koehler, 1996), in which the participants are presented with brief personality descriptions of several individuals, allegedly sampled at random from a group of e.g. 30 descriptions of engineers and 70 descriptions of lawyers. The participants are asked to indicate, for each description, their probability that the person described is an engineer. It is found (inter alia) that, when the description is totally uninformative with respect to the profession of the described individual, the median indicated probability is around 50% (rather than the presumably correct value of 30%). Gigerenzer and Murray (1987, p. 156)

[6] On the other hand, I agree with Gigerenzer (1991a, pp. 93–94) that statistical norms do not apply to the experimental context when the experimenter carelessly neglects to stipulate random sampling.

comment:

> consider the additional information in the Engineer-Lawyer problem that a panel of experts were "highly accurate" in the same task and that "you will be paid a bonus to the extent that your estimates come close to those of the expert panel"… The subjects may understand from the success of the experts that the personality descriptions are highly informative if only one knows how to read them. Thus they might conclude that there is only one strategy to win the bonus, namely, to concentrate on the description and to forget about the base rates.

The point seems to be that the application to E/L experiments of a general conversational norm which instructs listeners to interpret speakers' contributions to conversations as relevant (cf. Grice, 1989; Sperber & Wilson, 1995) creates a context for which statistical norms are inappropriate. (Schwarz, 1994; 1996 elaborates the idea that several biases are partly due to the logic of conversation).[7]

The above reasoning is inconclusive. Let me grant that in the above context conversational norms *conflict* with statistical norms: the former prescribe taking into account a description which is in fact uninformative, whereas the latter prescribe discarding the description. Still, Gigerenzer has not shown that one ought to disregard the statistical rather than the conversational norms; thus Gigerenzer is not entitled to conclude that statistical norms are inappropriate. In short, Gigerenzer has at most provided an *explanation*, not a *justification*, for the participants' disregard of statistical norms. (Note also that Gigerenzer's explanation seems inapplicable to other biases like overconfidence or even to cases in E/L experiments in which participants are presented with genuinely relevant information, namely informative descriptions.)

Gigerenzer (pers. commun.) has replied that he never intended to deny the appropriateness of statistical norms for experimental contexts; he intended rather to criticize the "content-blind" application of statistical norms to particular contexts (experimental or not). I agree with the latter criticism, but if this is Gigerenzer's whole point then I don't see the relevance of his appeal to conversational norms: there may indeed be reasons why the participants' neglect of base rates in E/L experiments is not irrational (as Mueser, Cowan and Mueser (1999) forcefully argued), but as I explained above the fact that attending to the base rates conflicts with conversational norms is not in itself such a reason.

4.2. Do conflicting statistical norms exist?

Gigerenzer and Murray (1987, p. 168) claim that "at the concrete experimental

[7] Gigerenzer and Murray also claim that almost all participants in E/L experiments know it is a lie that the descriptions with which they are presented were *randomly* chosen (Gigerenzer & Murray, 1987, p. 167; cf. Gigerenzer et al., 1989, p. 231). Even if this claim is true, it does not follow that the participants "must disregard the base rates" (Gigerenzer & Murray, 1987, p. 167), because the participants presumably have no clue about the experimenter's *precise reason* for lying to them and thus have no reason to disregard the base rates.

level statistics does not speak with one voice". To support this claim, they examine in detail (pp. 168–173; cf. pp. 157–162; Gigerenzer, 1991a, p. 104; Gigerenzer et al., 1989, pp. 228–231) a version of what is known as the "cab problem":

> A cab was involved in a hit-and-run accident at night. … 85% of the cabs in the city are Green and 15% are Blue. … A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases. … What is the probability that the cab involved in the accident was Blue rather than Green? (Tversky & Kahneman, 1980, p. 62).

To compute the "correct" answer by means of Bayes' theorem, it is necessary to assume that the reliability of the witness is the same under the testing conditions and under the conditions prevailing at the night of the accident. Birnbaum (1983) has noted that this assumption is unwarranted, because signal detection theory (backed by data) implies that reliability depends on base rates, and the base rates under the testing conditions differ from the base rates at the night of the accident. Different theories of how the witness operates lead to different answers to Tversky and Kahneman's question. Contrary to Gigerenzer and Murray, however, Birnbaum does not conclude that conflicting statistical norms apply to the cab problem; he concludes rather that "theories of signal detection and judgment are required to generalize from the court's test of the witness to the performance in the street" (Birnbaum, 1983, p. 91). Moreover, one might try to avoid the difficulty by carefully modifying the formulation of the cab problem:

> The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time (Tversky & Kahneman, 1982, p. 156).

Gigerenzer and Murray, however, point out that a difficulty may still exist:

> because the witness testified Blue, the testimony – if wrong – can only be a false alarm [i.e. mistakenly testifying Blue], not a miss [i.e. mistakenly testifying Green]. If the witness wants to avoid giving the impression that his testimony could be wrong, he may shift the criterion in the court's test so as to reduce false alarms at the cost of misses (Gigerenzer & Murray, 1987, p. 173).

It seems indeed that statistics will provide a correct answer to the cab problem only if one justifies the assumption that the witness did not shift his criterion. But this conclusion only underscores Gigerenzer's remark that care is needed to check one's assumptions when applying statistical norms (see previous subsection); we have yet to be shown that conflicting statistical norms exist.

It might seem that belief in the existence of conflicting statistical norms is widespread. For example, Cosmides and Tooby (1996, pp. 2–3) claim:

> professional probabilists themselves disagree – often violently – about the central issues in their field. Different probabilists looking at the same problem will frequently give completely different answers to it – that is, they will make contradictory claims about which answer is normatively correct.

Actually, however, from Cosmides and Tooby's claim it does not follow that professional probabilists accept conflicting norms; maybe they accept the same norms but disagree on how to *apply* them to particular problems. In any case, for the sake of argument let us assume that professional probabilists accept conflicting norms and see what follows. I wish to present two reasons why the consequences may not be so dramatic as one might think at first sight. First, even if competing statistical theories exist, these theories may still give the same answer for *particular* problems, maybe even for *many* particular problems. (Indeed, it is hard to see why anyone would take seriously a theory that would give weird answers to bread-and-butter problems.) Second, the existence of disagreement about which norm is appropriate does not imply that there is no fact of the matter about which norm is appropriate: maybe exactly one of the conflicting norms is. As an analogy, from the existence of disagreement about the existence of black holes it does not follow that there is no fact of the matter about whether black holes exist.

5. Conclusion

The topic of appropriate norms for judgments under uncertainty is of central importance to the psychology of reasoning but has been relatively neglected in the psychological literature. This paper represents an attempt to start filling this gap.

Gigerenzer complains that the heuristics and biases literature has somewhat uncritically assumed that norms appropriate for single-case judgments exist and are furnished by probability and statistics. To raise the heuristics and biases researchers from their "dogmatic slumber," Gigerenzer has expressed some reservations about the appropriateness of calling "errors" or "fallacies" some of the biases identified by these researchers. In this paper I took Gigerenzer's reservations seriously and I argued that some of them can be countered. First, do norms appropriate for single-case judgments exist at all? Gigerenzer has been misunderstood as claiming that they don't, whereas his real argument is that the predominance of frequentism places the burden of proof on those who claim that such norms exist. In response I denied that the predominance of frequentism generates such a burden of proof: I argued that (a) frequentism provides no good reason to doubt the meaningfulness of single-case probabilities and that (b) even if single-case probabilities were meaningless it would not follow that no norms appropriate for single-case judgments exist. Second, are probabilistic norms appropriate? I outlined three research programs aimed at showing that degrees of confidence ought to satisfy the axioms of probability. Third, are statistical norms appropriate? I agreed with Gigerenzer that when applying statistical norms one should carefully check the applicability of the assumptions that are standardly made to solve textbook problems. On the other hand, Gigerenzer's attack on statistical norms by appealing to conversational norms I found inconclusive: Gigerenzer is right that statistical norms can conflict with conversational ones, but this does not show that it is the statistical rather than the conversational norms which should be disregarded.

Where does this leave us? I think that Gigerenzer is right to stress that researchers working in the psychology of reasoning would do well to justify rather than uncritically assume the existence and the nature of the norms with which they compare the performance of the participants in their experiments. But contrary to Gigerenzer I presented reasons for believing that the result of this justification process will, after all, to a large extent vindicate probabilistic and statistical norms.

Acknowledgements

I wish to thank Gerd Gigerenzer, James Joyce, Daniel Kahneman, Richard Nisbett, Peter Railton, Norbert Schwarz, Stephen Stich, and some anonymous reviewers for helpful comments.

References

Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases* (pp. 294–305). New York: Cambridge University Press.

Birnbaum, M. H. (1983). Base rates in Bayesian inference: signal detection analysis of the cab problem. *American Journal of Psychology*, *96*, 85–94.

Carnap, R. (1945). The two concepts of probability. *Philosophy and Phenomenological Research*, *5*, 513–532.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*, 1–73.

Dudley, R. M. (1989). *Real analysis and probability*. Pacific Grove, CA: Brooks/Cole.

Finetti, B. de (1980). Foresight: its logical laws, its subjective sources. In H. E. Kyburg Jr., & H. E. Smokler (Eds.), *Studies in subjective probability* (2nd ed., pp. 53–118). Huntington, NY: Krieger. (Original work published 1937)

Gigerenzer, G. (1991a). How to make cognitive illusions disappear: beyond "heuristics and biases". In W. Stroebe, & M. Hewstone (Eds.), *European review of social psychology* (Vol. 2, pp. 83–115). Chichester, UK: Wiley.

Gigerenzer, G. (1991b). On cognitive illusions and rationality. In E. Eells, & T. Maruszewski (Eds.), *Probability and rationality: studies on L. Jonathan Cohen's philosophy of science (Poznán Studies in the Philosophy of the Sciences and the Humanities)* (Vol. 21, pp. 225–249). Amsterdam: Rodopi.

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In K. I. Manktelow, & D. E. Over (Eds.), *Rationality: psychological and philosophical perspectives* (pp. 284–313). London: Routledge.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright, & P. Ayton (Eds.), *Subjective probability* (pp. 129–161). Chichester, UK: Wiley.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: a reply to Kahneman and Tversky (1996). *Psychological Review*, *103*, 592–596.

Gigerenzer, G. (1998). Ecological intelligence: an adaptation for frequencies. In D. D. Cummins, & C. Allen (Eds.), *The evolution of mind* (pp. 9–29). New York: Oxford University Press.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: the use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 513–525.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review*, *102*, 684–704.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance: how probability changed science and everyday life*. Cambridge, UK: Cambridge University Press.

Gigerenzer, G., Todd, P. M., & ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Good, I. J. (1959). Kinds of probability. *Science*, *129*, 443–447.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: how intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*, 275–305.

Horwich, P. (1982). *Probability and evidence*. New York: Cambridge University Press.

Howson, C. (1995). Theories of probability. *British Journal for the Philosophy of Science*, *46*, 1–32.

Howson, C., & Urbach, P. (1993). *Scientific reasoning: the Bayesian approach* (2nd ed.). Chicago, IL: Open Court.

Joyce, J. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, *65*, 575–603.

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

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*, 582–591.

Keren, G. (1991). Calibration and probability judgments: conceptual and methodological issues. *Acta Psychologica*, *77*, 217–273.

Koehler, J. J. (1996). The base rate fallacy reconsidered: descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, *19*, 1–17.

Lewis, D. K. (1986). *On the plurality of worlds*. New York: Blackwell.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: the state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases* (pp. 306–334). New York: Cambridge University Press.

Mackie, J. L. (1973). *Truth, probability and paradox: studies in philosophical logic*. Oxford, UK: Clarendon.

Maher, P. (1993). *Betting on theories*. New York: Cambridge University Press.

Maher, P. (1997). Depragmatized Dutch book arguments. *Philosophy of Science*, *64*, 291–305.

McClennen, E. F. (1990). *Rationality and dynamic choice: foundational explorations*. New York: Cambridge University Press.

Mueser, P. R., Cowan, N., & Mueser, K. T. (1999). A generalized signal detection model to predict rational variation in base rate use. *Cognition*, *69*, 267–312.

Nisbett, R. E., & Ross, L. (1980). *Human inference: strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Piattelli-Palmarini, M. (1994). *Inevitable illusions: how mistakes of reason rule our minds*. New York: Wiley.

Ramsey, F. P. (1980). Truth and probability. In H. E. Kyburg Jr., & H. E. Smokler (Eds.), *Studies in subjective probability* (2nd ed., pp. 23–52). Huntington, NY: Krieger. (Original work published 1926)

Samuels, R., Stich, S., & Bishop, M. (in press). Ending the rationality wars: how to make disputes about human rationality disappear. In R. Elio (Ed.), *Common sense, reasoning and rationality (Vancouver Studies in Cognitive Science)* (Vol. 11). Oxford, UK: Oxford University Press.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Schwarz, N. (1994). Judgment in a social context: biases, shortcomings, and the logic of conversation. In M. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 26, pp. 123–162). San Diego, CA: Academic Press.

Schwarz, N. (1996). *Cognition and communication: judgmental biases, research methods and the logic of conversation*. Hillsdale, NJ: Erlbaum.

Sperber, D., & Wilson, D. (1995). *Relevance: communication and cognition* (2nd ed.). Oxford, UK: Blackwell.

Stein, E. (1996). *Without good reason: the rationality debate in philosophy and cognitive science*. Oxford, UK: Clarendon.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*, 1124–1131.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (pp. 49–72). Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases* (pp. 153–160). New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293–315.

von Mises, R. (1957). *Probability, statistics, and truth*. New York: Macmillan. (Original work published 1928)