Published in: *Behavioral and Brain Sciences, 12* (4), 1989, 764–765.

A General Algorithm For Pattern Recognition?

Gerd Gigerenzer
Department of Psychology, University of Constance, Germany

How does the brain make sense of the world? "In the same way that scientists do, and with the same tools," answer an increasing number of cognitive psychologists. The eighteenth-century mathematicians Laplace and Condorcet used their "probability of causes" to model the way scientists reason (Daston, 1988); Dominic Massaro now proposes the identical formula as an algorithm for pattern recognition in general and speech perception in particular. To show the generality of the algorithm (the "fuzzy logical model of perception [FLMP]") is the ambitious goal of Massaro's book: "Well-learned patterns are recognized in accordance with a general algorithm, regardless of the modality or particular nature of the patterns." (p. 16) The project of reducing pattern recognition to any algorithm, much less a single one, may strike many as overly ambitious. For the purposes of this commentary, however, I will accept Massaro's goal as in principle attainable and will try to invite him to clarify one of his major arguments by revealing its conceptual difficulties.

This argument is central to the issue of generality, and runs like this: (i) The FLMP is not only general but also optimal since it is mathematically equivalent to Bayes' theorem; (ii) here Bayes' theorem is implemented as it was by Laplace, that is, with the assumption of uniform prior probabilities and independence of events (features), but (iii) this equivalence poses a dilemma for the FLMP, since previous research, in particular on intuitive probabilistic reasoning, has rejected Bayesian reasoning as a general mental algorithm. In analyzing this argument, I shall proceed from the general to the specific.

Is the General Algorithm a Bayesian One?

The FLMP assumes that pattern recognition occurs in three sequential stages. I shall consider the simplest case, with only two features and two prototypes. In the feature evaluation stage, the match $t(E_1/H_1)$ between a feature E_1 and a prototype H_1 is calculated; in the feature integration stage the overall match $t(E/H_1) = t(E_1/H_1)t(E_2/H_1)$ between the two features and the prototype H_1 is calculated; and in the pattern classification stage the probability that the pattern will be identified as H_1 is given by $p(H_1/E) = t(E/H_1)/(t(E/H_1) + t(E/H_2))$. Massaro (pp. 196–198) says that this algorithm is mathematically equivalent to Bayes' theorem, assuming uniform prior probabilities and independent events, and replaces the above t (i.e., truth) values by p (i.e., probability) values. However, Massaro repeatedly (e.g., pp. 21, 166, 202) asserts that $t(E_1/H_1) + t(E_1/H_2) = 1$, which is not true for the corresponding probabilities $p(E_1/H_1)$ and $p(E_1/H_2)$ in Bayes' theorem. According to standard probability theory, which mathematically implies Bayes' theorem, the sum of these probabilities can be either less or more than 1. Thus, I doubt that the proposed general algorithm is in fact mathematically equivalent to Bayesian probabilities, and I therefore also doubt the claim that "either of these two models is adequate to account for the results" (p. 198).

A General Pattern Recognition Algorithm With Uniform Priors?

To keep the next points separate from the first, let me assume that I have overlooked something and that Massaro is right in pointing to the equivalence of the FLMP and Bayes' theorem. Laplace's urn analogy and Bayes' billiard table suggested uniform prior probabilities on the grounds that our ignorance gives us no reason to expect one urn or one area on the table to be a priori more likely than any other. But should we assume that a general pattern recognition algorithm also works on the principle of ignorance and uses uniform priors? An algorithm with uniform priors may be sufficient for the experimental designs reported in the book, in which the prototypes to be identified, such as /ba/ and /da/, are equally likely in the laboratory. But in everyday speech, just as in many other domains, different patterns have different prior probabilities depending on context. Where expectation plays a role, nonuniform priors seem to be indispensable for improving the perceptual "bet" in situations with uncertain information.[1]

Does Intuitive Probabilistic Reasoning Challenge the Generality of the Algorithm?

Massaro argues that the mathematical equivalence of the FLMP and Bayes' theorem "poses a new dilemma" for the generality claim, since previous research has rejected Bayes' theorem as a model of intuitive probabilistic reasoning. His major defense is that most previous researchers used "objective" rather than "subjective" probabilities to calculate the so-called normative Bayesian outcome. Massaro's reply is correct: There are many ways to be a Bayesian, and such experiments do not rule out that intuitive reasoning is Bayesian by using subjective probabilities. But there are designs, such as in Kahneman and Tversky's (1973) Engineer-Lawyer study, which allow for subjective likelihoods and which still lead the authors to conclude that reasoning is not Bayesian, since base rates are ignored due to a representativeness heuristic. Massaro has to deal with these kinds of experiments. Moreover, even in the Engineer-Lawyer problem, the neglect of base rates can easily be eliminated if one crucial structural assumption (random sampling of description) is made vivid to subjects, although the feature values remain constant (Gigerenzer et al., 1988). Such systematic changes in reasoning indicate that neither representativeness (i.e., uniform-prior Bayesianism, see below) nor Bayes' theorem is a general algorithm of the mind. Massaro's "dilemma," in my opinion, is not that intuitive reasoning ignores base rates (as does the FLMP in the book under review), but rather that intuition seems to have a whole toolbox of algorithms available.

Two things puzzle me concerning the relation between reasoning and the FLMP. First, as noted above, Massaro says that the pattern recognition algorithm is a Laplacean uniform-prior variant of Bayes' theorem. Why then does he believe it is a "dilemma" that intuitive reasoning seems to violate Bayesian reasoning by neglecting base rates and using uniform priors? Base rate neglect is exactly what his algorithm predicts—just as Baconian probability would (Cohen, 1986). [See

---

1 However, nonuniform priors may not be required to improve the fit of the FLMP to experimental data, since it is already excellent. My point here is a conceptual one. It is in fact hard to judge to what degree the FLMP is supported by the empirical data, because of the large number of free parameters (sometimes close to 50% of the data points) that can be fitted to the data. The true degree of support could be revealed by a step-wise cross-validation procedure: Use the best-fitting feature values for given prototypes and a given subject in a new experiment with the same or an enlarged set of features. If features are evaluated independently from other features present, as the FLMP proposes, then the feature values should be stable, and the fit in the second experiment would provide a stronger test for the validity of the FLMP.

also Cohen (1981), Can Human Irrationality Be Experimentally Demonstrated?, *BBS, 4* (3) and Kyburg (1983), Rational Belief, *BBS, 6* (2).]

Second, what is the relationship between Kahneman and Tversky's representativeness heuristic and the FLMP as models of pattern recognition? Massaro says they are fundamentally different (p. 273). However, since in this context "representativeness" can be shown to mean Bayesian reasoning with uniform priors (Gigerenzer & Murray, 1987, chap. 5), I understand both the FLMP and the heuristic to refer to the same strategy—although the FLMP is a model and representativeness is just a word.

The "Conjunction Fallacy" and the FLMP

What has been called a "conjunction fallacy" is a judgment of the following kind: A fictitious person named Linda is more likely to be a feminist and a bank teller than just a bank teller. I agree with Massaro that this cannot be called a "fallacy" unless one willfully ignores the fact that the term "likely" has several meanings in everyday language. But I part ways with him when he extends the generality of the FLMP to conjunction judgments and claims that "within Bayes theorem or the FLMP, we can predict that Linda will be rated as being more likely [to be] a feminist and bank teller than just a bank teller" (p. 275). Massaro's argument is that the subjects behave as if they were carrying out pattern recognition, evaluating the features against alternative prototypes. In fact, the FLMP predicts that the probability that a pattern is recognized as prototype H_1 can be larger for two features E_1 and E_2 than for just E_1 alone. In formal terms, this means that $p(H_1/E_1\&E_2) > p(H_1/E_1)$. But this is not the "conjunction fallacy," which is to judge $p(H_1\&H_2/E_1) > p(H_1/E_1)$. Only the latter contradicts standard probability theory, and thus cannot be derived from either Bayes' theorem, which is a consequence of standard probability theory, or the FLMP, insofar as it is claimed to be mathematically identical to the former.[2]

Why only one algorithm?

Scientific reasoning is a many-splendored thing, encompassing many and diverse forms of inference. We might want to carry the analogy between scientific reasoning and cognition far enough to extend this multiplicity to the mind. Thus, instead of one general, all-purpose algorithm, we would expect to find several or even many algorithms. There is an additional Darwinian argument against the view that evaluation has given us only a single algorithm for all cases of pattern recognition. Since each algorithm assumes a specific structure (e.g., in Massaro's model, independence of features and a mutually exclusive and exhaustive set of prototypes), it is well equipped for specific tasks that have these structures, but less so for other tasks. In order to survive in a changing environment, the mind would be better off if outfitted with a whole toolbox of algorithms, and with an evaluation program that first checks the structure of the environment before it selects a particular algorithm to apply to that environment.

---

2 In his own experiments, Massaro (submitted) showed that the FLMP gives an excellent fit to judgments in Linda-type tasks. However, judgments of the type *p*(Linda/vocation&avocation) were compared with *p*(Linda/vocation) and *p*(Linda/avocation), which, as mentioned above, cannot violate the conjunction rule.

I have concentrated in this commentary on conceptual issues that need further clarification. These have to do with the claims for a general algorithm for pattern recognition, but do not touch the valuable experimental work that the book presents, which I have not here addressed.