The Validity of
Dempster–Shafer
Belief Functions

Gregory M. Provan

Department of Computer and Information Science
University of Pennsylvania
Philadelphia, Pennsylvania

ABSTRACT

This reply to papers by Pearl and Shafer focuses on two issues underlying the debate on the validity of using Dempster–Shafer theory, namely the requirement of a process-independent semantics and the a priori need for multiple uncertainty calculi. Pearl shows deficiencies of Dempster–Shafer theory in dealing with several instances of commonsense reasoning in a process-independent manner. Although this argument is correct under the assumptions stated, it is weakened somewhat by introducing questions of whether a process-independent semantics is always necessary or desirable.

Another issue underlying both papers, whether multiple uncertainty representations are necessary, is also discussed. Shafer claims that multiple uncertainty representations are necessary. He presents a goal of developing all uncertainty representations in parallel and defining domains in which each representation is best suited. In contrast, Pearl implicitly claims that probability theory alone is necessary, unless the use of another representation (such as Dempster–Shafer theory) is shown to be clearly advantageous. These two perspectives lead to different approaches to defining the form of uncertainty best modeled by Dempster–Shafer theory or any other uncertainty calculus.

KEYWORDS: *Dempster–Shafer theory, probability theory, uncertainty representations*

1. INTRODUCTION

Both papers by Pearl [1] and Shafer [2] attempt to more carefully delineate the domain of applicability of Dempster–Shafer theory (DST). Among other things, Shafer discusses the semantics of several interpretations of DST,

thereby providing insight into the appropriate applications of DST. Pearl shows that applying DST to represent tasks involving (1) incomplete knowledge, (2) belief updating, and (3) evidence pooling, tasks for which DST has been claimed appropriate, in fact either produces counterintuitive results or is otherwise inadequate.

It is important to delineate the domain of applicability of all uncertainty calculi. Probability theory is arguably the best understood uncertainty calculus, and DST needs further analysis. Clearly, analyses of DST as provided by Shafer and Pearl are needed. However, there seem to be underlying (and conflicting) motivations to the ways in which DST has been analyzed by Shafer and Pearl.

First, the two authors differ on whether a process-independent semantics is fundamental to reasoning tasks. This assumption makes a big difference to whether a calculus like DST is widely applicable. Second, they differ on the extent to which multiple uncertainty calculi are necessary and/or desirable. Shafer starts from the point of view that reasoning under uncertainty *requires* multiple calculi, each calculus being used for given tasks to which it is best suited. DST, it is claimed, is a calculus that is most appropriate for a set of tasks that are outlined in that paper [2]. In contrast, Pearl, a staunch defender of probability theory [3, 4] (and see the related arguments of Cheeseman [5, 6]), starts from the point of view that a single uncertainty calculus, probability theory, is sufficient, unless the superiority of some alternative calculus (over probability theory) can be proved precisely. From this perspective, Pearl aims to show that DST is inferior to probability theory for tasks for which DST has been claimed suitable, and hence cannot supplant probability theory for these tasks.

In this reply, issues relating to process-independent semantics and uncertainty calculi are discussed in Section 2. Questions concerning the need for multiple uncertainty calculi, and how DST fits into such a scheme of single/multiple calculi, are discussed in Section 3. A few conclusions are then presented.

2. SEMANTICAL ISSUES

In this section, the issue of whether a process-independent semantics is the best one for uncertain reasoning is discussed. A process-independent semantics is one in which the semantics of propositions is independent of how the propositions are derived. Such a semantics provides a powerful normative theory of belief ascription, among other things.

This argument focuses on whether such a semantics is appropriate for all reasoning tasks. One interpretation of DST is in terms of a process-*dependent* assignment of conditional probability to the provability of a proposition, given

that the evidence for the proposition is noncontradictory. Enforcing a process-*independent* semantics on DST renders it incorrect in certain circumstances (as shown by Pearl [1]), whereas the process independence of probability theory still produces correct results. However, if indeed the *means* of obtaining results is important, then DST will be correct for some of the cases for which Pearl shows it is incorrect.

To shed some light on these questions, possible formalizations of commonsense reasoning are introduced.

2.1. Possible Notions of Commonsense Reasoning

This section introduces questions of whether a process-independent semantics is always necessary or desirable. To determine what form of semantics is necessary, one needs to define commonsense reasoning (CSR). Pearl never precisely defines CSR; he is not to be faulted for this, as very few people who do research in CSR do either. CSR has to do with rather vague tasks like reasoning with exceptions, such as "birds fly," knowing that there exist flightless birds.

Questions to be posed include "What is an exception?", or "How are default rules deduced?" It is not clear whether these questions can be evaluated over time, for example, with trial and error, so that a relative frequency semantics (which is process-independent) can be assigned, or whether the process of determining the default assignment is important.

There are a multitude of possible interpretations of what CSR might be about. This review is meant to be illustrative, and not comprehensive.

RELATIVE FREQUENCY INTERPRETATION A relative frequency interpretation of a default rule is based on notions of frequency of events. It might be true that, over the process of evolution, certain "default rules" may have been established by trial and error and/or observation over large samples (i.e., many individuals over a large time span). For example, consider the situation proposed by Boden [7], who examines how a kingfisher might catch fish if it does not explicitly know Snell's law of light's refraction in water. She hints at Snell's rule being hardwired owing to the evolutionary experience of the kingfisher; that is, out of the population of kingfishers who dive in multiple fashions, the kingfishers that just happened to dive using Snell's law ate well and prospered, and all other kingfishers ate less well because they caught fewer fish owing to the refractive properties of water. Over time the average kingfishers were replaced by the healthier birds obeying Snell's law.

Aleliunas [8] presents a similar evolutionary argument. He argues that knowledge about the world is statistical knowledge (presumably compiled over the process of evolution). The strong claim that intelligence necessitates the use of these statistics is made using a Dutch book argument.

In a sense, then, a default rule can be seen to have been derived from a sample set, and relative frequency statistics can be used to analyze the default. This is one of many possible interpretations of defaults. For domains in which this interpretation is appropriate, a frequentist approach is best, and DST is relatively less appropriate.

However, it is unclear whether this frequentist interpretation of default rules could be true in all cases. In many situations, it appears that default rules are established and applied with relatively little experience. For example, biases are often established over a very small sample set but are applied to a large population. Buying a green apple that makes one sick may lead to a bias against non-red apples, even though Granny Smith and Golden Delicious apples are perfectly edible. In the apple example, modeling the default "Non-red apples are inedible" relies entirely on how the default was created, and not on statistical arguments about the proportion of green apples that are inedible.

ARGUMENT SYSTEMS A second possible interpretation of CSR is given by argument systems (Loui [9], Poole [10])—for each proposition believed true, an argument is constructed concerning why it should be believed. In this case the *structure* of the argument (e.g., whether it is the shortest argument) is of primary importance.

Shafer [2] suggests that procedural methods should be used. This entails that the reliance on declarative semantics may be throwing us off the track of modeling CSR and hence is undesirable: "process-independent semantics is *not* a reasonable goal for AI" as "judgments under uncertainty . . . depend on the process by which they were made as well as on the objective nature of the evidence." Shafer also cites Winograd [11] in claiming that AI must use statements that have no meaning "in any semantic system that fails to deal explicitly with the reasoning process." If this argument were adopted, the notion of provability underlying DST (Pearl [4], Provan [12]) would be more appropriate than the declarative semantics of probability theory. But this in turn necessitates a closer examination of what is meant by CSR.

One approach to studying CSR has been the analysis of belief ascription and revision. In particular, there has been a great deal of research on belief revision (e.g., Alchourron et al. [13], Shoham [14], Martins and Shapiro [15]) and on an implementation of belief revision, truth maintenance systems (TMSs) (Doyle [16], de Kleer [17], Martins and Shapiro [15]). A TMS records the arguments for the belief of propositions (in terms of assumptions or rules underlying the proposition). It is interesting to note that the semantical relationship between TMSs and DST has been used as a means of implementing DST (Laskey and Lehner [18], Provan [19]).

SUBJECTIVE PROBABILITY A subjective interpretation of probability can be assigned a semantics according to betting behavior, given the subjective

assignment of probabilities. Note that this semantics is process-independent. This is the probabilistic interpretation favored within AI and is the one championed by Pearl. One point not raised in Pearl's critique of the difficulties in establishing a coherent model for defaults using DST is that all models, including the subjective probability model [e.g., ϵ-semantics (Pearl [4])], have had difficulty.

Clearly, assigning a coherent semantics to defaults is a difficult task, and research could probably be furthered by examining the nature of what is to be modeled in a default and how a default is best used in the real world (and not just the Tweety/Opus flight domains). Without precisely defining default rules one cannot condemn an uncertainty calculus for not modeling a poorly defined paradigm.

2.2. Normative/Subjective Questions

Another question with CSR (and uncertainty calculi for CSR) is whether the goal is to simulate human CSR or to enable computers to reason in a normative manner (possibly given small samples of data). There are many reasons to use a purely normative reasoning system. However, the computational expense typically required by such systems entails the use of approximation techniques and/or heuristics. The subjective reasoning of humans may be a rich source for such techniques. For example, medical expert systems use rules and probabilities derived from experts' experience. Such probabilities are now combined in a normative manner [as opposed to the heuristic methods of earlier systems like MYCIN (Buchanan and Shortliffe [20])]. However, the subjective origins (and possible subsequent use for diagnosis and treatment) of the probabilities cannot be forgotten.

This is not necessarily an argument for subjectivism, but a reflection on the fact that purely normative methods may be infeasible in many circumstances.[1]

2.3. Questions of Utility

What is the CSR conclusion to be used for? All CSR judgments are utility-based. For example, if it is a life-and-death decision, then it is important that the decision be a good one. But if it has trivial consequences, for example, I can take either the 2:14 bus or 2:30 train to New York, and both will get me there around 3:30, then a poor decision will have little consequence.

The need for utility functions argues strongly in favor of probability theory, with its well-understood decision theory. DST has no notion of utility compa-

---

[1] This is similar to a conclusion reached in an analysis of the normative use of logic for reasoning (McDermott [21]), namely that the normative approach to reasoning has been partially successful and further research in both normative and subjective approaches is necessary.

rable to that of probability theory, although arguments have been made about ascribing such notions (Smets [22], Jaffray [23]). Further research is clearly necessary.

3. SINGLE-CALCULUS VS. PLURALIST VIEWPOINT

One of the fundamental issues underlying the two papers of Pearl [1] and Shafer [2] is whether there is a single uncertainty calculus appropriate for all forms of reasoning. It is desirable to have a single calculus as the foundation for reasoning from many perspectives; these issues are not discussed here. The question at issue here is whether that is possible.

Shafer adopts a pluralistic viewpoint, namely that probability theory, Dempster–Shafer theory, Fisherian and Neyman–Pearson methods, fuzzy set theory, etc., can all play vital roles in reasoning under uncertainty. He argues that "the mere fact that there is uncertainty in a problem does not mean that the theory of probability is useful in the problem." Shafer makes the guarded statement [2, Sec. 2] that modeling the uncertainty for a given problem must be done on a case-by-case basis and cannot be done simply by determining the calculus (probabilistic or otherwise) best suited to that general class of problem. Nonetheless, establishing the advantages and disadvantages of using specific calculi to model the uncertainty present in given classes of tasks must help in narrowing the types of calculi most suitable for a particular problem.

Shafer is attempting to more precisely define the role Dempster–Shafer theory can play in uncertain reasoning, given his world view of many uncertainty calculi playing various roles according to which representation is best for each role. In fact, implicit in Shafer's pluralistic view is the notion that *no* uncertainty calculus is optimal or adequate for all uncertainty reasoning tasks.

In contrast, Pearl argues from the viewpoint that probability theory is the *accepted* calculus for reasoning under uncertainty. Hence, for a new calculus X to be accepted, it must be shown either that probability theory is unable to model a certain form of uncertainty (while X can) or that X can model a form of uncertainty better than probability theory [e.g., X demonstrates greater efficiency, imposes fewer unwarranted requirements (such as variable independence), etc.] Historically, many alternative calculi have been proposed and implemented in systems without clear demonstration of their advantages over probability theory.[2] Hence, Pearl is forcing the proponents of alternative

---

[2] Both MYCIN certainty factors (Buchanan and Shortliffe [20]) and DST have been guilty of being used inappropriately.

calculi (like DST) to precisely define the domains for which the calculi are best suited and not merely assume that it is more advantageous to use a calculus with certain properties.

3.1. Origins of Pluralistic Viewpoint

Many different uncertainty calculi have been developed to circumvent the perceived problems of probability theory. These perceived inadequacies include the following.

- *Priors* One well-known problem is the large number of priors required. In cases where there are many unknown or missing priors, probability theory cannot be used effectively; it is for such problems that DST is claimed to be useful.
- *Psychological validity* The justification that probability theory is used by people in commonsense reasoning (CSR) is not validated by psychological investigation. Tversky and Kahneman [24] have shown that most people (even trained mathematicians) have difficulty using probability theory correctly. Several other studies (Tversky and Kahneman [25, 26], Birnbaum [27]) have shown that reasoning is done using some calculus other than probability theory. In examining the evidence, Smithson [28] concludes:

> Even after one accounts for the lack of specification in many of the problems and tasks, possible confounding or artifactual factors, and the potential for nonshared understanding between subject and experimenter, there is still sufficient residuum of evidence to suggest strongly that Bayesian (or any other kind of) probability does not fit what many people do when making judgments and decisions under uncertainty. The dominant trend in the literature has been to account for the subjects' judgements by fairly simple "heuristics."

3.2. Discussion

A question that arises is whether probability theory (or any other knowledge representation calculus) can or should be a basis for CSR. What is desired is a holy grail, the underlying cognitive calculus by which all forms of reasoning are dictated. Both logic (McCarthy and Hayes [29], Moore [30]) and probability theory (Cheeseman [5, 6]) have been proposed as the calculus for all reasoning. A detailed discussion of the arguments for and against the validity of such claims is not presented here. The main argument made here is that a single-calculus point of view can have an adverse affect upon scientific research. In particular, some reasons for studying multiple uncertainty calculi in parallel, rather than a single calculus, are presented.

First, the issue of whether probability can model all forms of uncertainty and ignorance is as yet unresolved. For example, it is claimed that the uncertainty (e.g., vagueness) modeled by fuzzy set theory cannot be modeled by probability theory (Zadeh [31, 32]); moreover, even within fuzzy set theory Klir [33] has proposed a separation of "vagueness" from "ambiguity." These issues are not discussed here for lack of space; Smithson [28] presents a detailed analysis of the many arguments. Smithson argues that these issues are unresolved, but it appears that multiple uncertainty calculi are indeed necessary, owing to the inability of any single calculus to model all types of uncertainty. Strict adherence to a single-calculus viewpoint means that some forms of uncertainty may not be modelable within that calculus.

A second issue is whether the same uncertainty calculus can be used for all domains, let alone for many types of uncertainty that can arise in any one domain. It may be the case that different uncertainties are required for different domains, such as machine vision, natural language, problem solving, and game playing.

Acceptance of a pluralism of uncertainty representations changes the way in which DST (or any other uncertainty calculus) is studied. A rigid single-calculus approach may attempt to undermine the validity of DST through (possibly unfair) comparisons within a setting inappropriate to its application. For example, consider two incorrect interpretations of DST in terms of probability theory:

- The interpretation of DST belief and plausibility measures as probability bounds, which is inconsistent with Dempster's combination rule (Zadeh [34]); or
- The process-independent interpretation of DST mass functions and belief measures, which can lead to counterintuitive results (Pearl [1]).

In contrast, the pluralist approach attempts to determine how best DST can be used, in settings where its semantics are valid. It may turn out that DST (and many recently devised uncertainty calculi) are inferior to more established calculi (e.g., probability theory) for a variety of reasons, such as computational efficiency or lack of a domain that the calculus models in a clearly superior manner. If that occurs, such calculi will be abandoned. But these new calculi must be judged fairly and evaluated using conditions in which they would actually be appropriate. However, if one adopts a pluralist approach, parallel exploration of multiple uncertainty calculi must be conducted using the same precise mathematical criteria with which probability theory has been analyzed. It is clear that the sloppy analyses that have been applied to DST, certainty factors, and other alternative calculi are inadequate and indeed counterproductive. What is needed is a precise analysis based on criteria relevant to the semantics of each calculus. Note that this proposal does not rule out the extension of any single calculus to model forms of uncertainty previously not modeled by that calculus; for example, consider the extension of

probability theory to include the exploration of second-order probabilities (Pearl [35], Kyburg [36]).

4. CONCLUSIONS

If Pearl's assumptions [1] are correct, we agree with his demonstration of the deficiencies of DST in dealing with several instances of CSR in a process-independent manner. However, his argument is weakened somewhat by questioning whether a process-independent semantics is always necessary or desirable.

Another issue underlying both papers, the utility of studying multiple uncertainty representations in parallel, is also discussed. Shafer [2] claims that multiple uncertainty representations are necessary. Therefore, the goal is to develop all uncertainty representations in parallel and define domains in which each representation is best suited. We argue that the goal of justifying probability theory, or any other uncertainty calculus, as the *only* uncertainty calculus is a counterproductive research agenda, as it may lead to an inability to model all forms of uncertainty. However, tacit acceptance and study of new uncertainty calculi without precise analysis is also counterproductive. Since probability theory is the best developed calculus to date, the burden is thus placed on the proponents of DST and other newer uncertainty calculi to precisely define the semantics, computational properties, and domains of applicability of these theories, thereby showing the superiority of such calculi with respect to probability theory.

References

1. Pearl, J., Reasoning with belief functions: an analysis of compatibility, *Int. J. Approx. Reasoning* **4**(5/6), 363–390, 1991.

2. Shafer, G., Perspectives on the theory and practice of belief functions, *Int. J. Approx. Reasoning* **4**(5/6), 323–362, 1991.

3. Pearl, J., How to do with probabilities what people say you can't, *Proceedings of 2nd Conference on AI Applications*, Miami, Fla., 1985.

4. Pearl, J., *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, Los Altos, Calif., 1988.

5. Cheeseman, P., An enquiry into computer understanding, *Comput. Intell.* **1988**, 58–68.

6. Cheeseman, P., In defense of probability, in *Proceedings of the International Joint Conference on Artificial Intelligence* **1985**, 1002–1009.

7. Boden, M. A., Animal perception from an artificial intelligence viewpoint, in *Minds, Machines and Evolution* (C. Hookwag, Ed.), Cambridge Univ. Press, New York, 1984, pp. 153–174.

8. Aleliunas, R., Intelligent creatures must be statisticians, in *Proceedings Symp. Internacional de Inteligencia Artificial*, 21–36, Monterrey, Mexico, 1988.

9. Loui, K. P., Defeasible decisions: what the proposal is and isn't, in *Uncertainty in Artificial Intelligence* (L. Kanal, M. Henrion, R. Shachter, and Lemmer, J., Eds.), Elsevier, North Holland, Amsterdam, 1990, pp. 99–116.

10. Poole, D. L., On the comparison of theories: preferring the most specific explanation, *Proceedings of the International Joint Conference on Artificial Intelligence*, 144–147, 1985.

11. Winograd, T., Extended inference modes in reasoning by computer systems, *Artificial Intelligence* 13, 5–26, 1980.

12. Provan, G., A logic-based analysis of Dempster–Shafer theory, *Int. J. Approx. Reasoning* (Special issue on belief functions and belief maintenance in AI), 4, 451–498, 1990.

13. Alchourron, C., Gardenfors, P., and Makinson, D., On the logic of theory change: partial meet contraction and revision functions, *J. Symbolic Logic* 50, 510–530, 1985.

14. Shoham, Y., *Reasoning About Change*, Cambridge Univ. Press, New York, 1988.

15. Martins, J., and Shapiro, S., A model for belief revision, *Artificial Intelligence* 25-80, 1988.

16. Doyle, J., A truth maintenance system, *Artificial Intelligence* 12, 231–272, 1979.

17. de Kleer, J., An assumption-based TMS, extending the ATMS—problem solving with the ATMS, *Artificial Intelligence* 28:127–224, 1986.

18. Laskey, K. B., and Lehner, P. E., Assumptions, beliefs and probabilities, *Artificial Intelligence* 41, 65–77, 1990.

19. Provan, G., An analysis of ATMS-based techniques for computing Dempster–Shafer belief functions, *Proceedings of the International Joint Conference on Artificial Intelligence* 1115–1120, 1989.

20. Buchanan, B. G., and Shortliffe, E. H., *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Boston, 1984.

21. McDermott, D. V., A critique of pure reason, *Comput. Intell. J.* 3, 151–160, 1987.

22. Smets, P., Constructing the pignistic probability function in a context of uncertainty, in *Uncertainty in Artificial Intelligence* (L. Kanal, M. Henrion, R. Shachter, and Lemmer, J., Eds.), North-Holland, Amsterdam, 1990, pp. 29–39.

23. Jaffray, J., Linear utility for belief functions, *Op. Res. Lett.* 8, 107–112, 1989.

24. Tversky, A., and Kahneman, D., Judgement under uncertainty: heuristics and biases, *Science* **185**, 1124–1131, 1974.

25. Tversky, A., and Kahneman, D., Causal schemas and judgements under uncertainty, in *Progress in Social Psychology*, Vol. I (M. Fishbein, Ed.), Erlbaum, Hillsdale, N.J., 1980.

26. Tversky, A., and Kahneman, D., Extensional versus intuitive reasoning: the conjunction fallacy in probability judgement, *Psychol. Rev.* **90**, 293–315, 1983.

27. Birnbaum, M. H., Base-rates in Bayesian inference, *Am. J. Psychol.* **96**, 89–54, 1980.

28. Smithson, M., *Ignorance and Uncertainty: Emerging Paradigms*, Springer-Verlag, New York, 1989.

29. McCarthy, J., and Hayes, P., Some philosophical problems from the standpoint of artificial intelligence, in *Machine Intelligence* (B. Meltzer and D. Michie, Eds.), American Elsevier, New York, 1969, pp. 463–502.

30. Moore, R. C., The role of logic in knowledge representation and commonsense reasoning, *Proceedings of the American Association for AI*, 428–433, 1982.

31. Zadeh, L., Probability measures of fuzzy events, *J. Math. Anal. Appl.*, **23**, 421–428, 1968.

32. Zadeh, L., The role of fuzzy logic in the management of uncertainty in expert systems, *Fuzzy Sets Syst.* 199, 1983.

33. Klir, G., Where do we stand on measures on uncertainty, ambiguity fuzziness and the like?, *Fuzzy Sets Syst.* **24**, 141–160, 1987.

34. Zadeh, L., A simple view of the Dempster–Shafer theory of evidence, and its implications for the rule of combination, *AI Mag.* **7**, 85–90, 1986.

35. Pearl, J., Do we need higher-order probabilities, and, if so, what do they mean?, *Proceedings Conf. Uncertainty in AI*, 47–60, 1987.

36. Kyburg, H., Higher order probabilities, in *Proceedings of the Conference Uncertainty in AI*, 22–29, 1987.