Comments on (1) Shepard; (2) Tenenbaum and Griffiths; and (3) Barlow.
(Abstract: 69; Main Text: 1639: Reference: 203; Total: 1912.)

"Universal Bayesian Inference": Some Comments on Shepard, and Tenenbaum and Griffiths

David Dowe

Department of Computer Science, Monash University, Clayton, Vic 3800, Australia.

David.Dowe@infotech.monash.edu.au

http://www.cs.monash.edu.au/~dld

Graham Oppy

Department of Philosophy, Monash University, Clayton, Vic 3800, Australia.

Graham.Oppy@arts.monash.edu.au

We criticise Shepard's notions of 'invariance' and 'universality', and the incorporation of Shepard's work on inference into the general framework of his paper. We then criticise Tenenbaum and Griffiths' account of Shepard (1987), including the attributed likelihood function, and the assumption of 'weak sampling'. Finally, we endorse Barlow's suggestion that MML theory has useful things to say about the Bayesian inference problems discussed by Shepard, and Tenenbaum and Griffiths.

Shepard (2001) claims that it is a general fact about the world that objects which are of the same basic kind generally form a connected local region in the space of possible objects. *Prima facie*, at least, this is not a 'fact about the world' at all; rather, it is an analytic or *a priori* truth which connects together the notions of 'basic kind' and 'connected local region in the space of possible objects'. Moreover, even if this were a general fact about the world, it seems implausible to suppose that it would have an important role to play in the explanation of the 'universality' of Bayesian inference: the efficacy of Bayesian inference does not depend upon the <u>kinds</u> of things which there are. The 'universality, invariance and mathematical elegance' which Shepard finds for this theory of inference seems different in kind from that which he finds for his theory of perceived colour—and perhaps also for his theory of perceived motion—even though it might be accommodated in an appropriate evolutionary theory (on the grounds that Bayesian inferrers will be advantaged in any possible world in which mobile perceivers can evolve).

Shepard's use of the term 'invariance'—in characterising the general aim of this theory—puzzles us. He claims that the General Theory of Relativity provides a model for the kind of psychological theory which he wants. However, it seems to us that there are several confusions in this part of his discussion. Most importantly, while it is true that Einstein cast General Relativity in generally covariant form, it is perfectly possible to cast Newtonian mechanics in generally covariant form: this just amounts to the observation that it is possible to give a coordinate–free formulation of these theories. In our view, the most natural notion of invariance—or invariance group—for physical theories is that developed by Anderson (1967), which turns on questions about the 'absolute objects' postulated by these theories: the advance which is marked by General Relativity is that it contains no absolute objects. Furthermore, as we understand it, it was never the case that Newton's laws were restricted to inertial frames moving at slow speeds relative to the speed of light; rather, it turns out that Newton's 'laws' do not fit the data for objects moving at speeds which are not slow relative to the speed of light. We confess that we have no idea what Shepard has in mind when he says that he is aiming for 'invariant' psychological principles. (We have similar concerns about his use of the term 'universal', and its relations to 'invariance'.)

Tenenbaum and Griffiths (2001) begin with 'Shepard's problem of generalisation from a single positive instance'. They propose a modification—to what they take to be Shepard's approach—which makes a significant difference when we move on to consider generalisation from multiple positive instances. In their view (Section 2.3), Shepard (1987) argued for the default assumption that the positive instance was sampled uniformly from the full range of cases ('weak sampling'); Tenenbaum and Griffiths claim that in many cases, it is better to suppose that the positive instance is sampled uniformly from the consequential region, i.e., from amongst the positive cases ('strong sampling'). Moreover, Tenenbaum and Griffiths suppose that this difference shows up as a difference in the likelihood functions for the two approaches. They favour the likelihood function which takes the value $1/|h|$ when $x \in h$ and is otherwise 0; and they attribute to Shepard the likelihood function which takes the value 1 when $x \in h$, and is otherwise 0. We think that something has gone wrong here: the likelihood function attributed to Shepard has not been normalised, and the fact that it attributes probability 0 to data outside the consequential region contradicts the independence in their informal definition of 'weak sampling'. Moreover, this function makes no sense as a value for the conditional probability of observing x given that h is the true consequential region: under the assumption that the sampling is uniform from the entire range of cases, that value would have to be 1/| C|, where C is the measure of the entire space. Perhaps Tenenbaum and Griffiths have here conflated the likelihood function with the probability that the sample lies in the consequential region (which is indeed 1 if the sample is taken from the consequential region).

There may be reason for caution in attributing 'weak sampling' to Shepard. It seems that Shepard does make this assumption (1987:1321)—though we didn't find Shepard's text entirely clear, and we have some sympathy for Tenenbaum and Griffith's interpretation of it—but only in the context of his 'theoretical justification' for the choice of a probability density function for the *size* of the consequential region. What he says is that, *in the absence of any information to the contrary*, an individual might best assume that nature selects the consequential region and the first stimulus independently. However, in the cases which Tenenbaum and Griffiths consider, it is plausible that Shepard would deny that there is no further information to the contrary—the baby robin has reason to suppose that mother will have sampled from the consequential region; and the doctor knows that the patient has been sampled from the consequential region—so it is not clear that Shepard can't get the same answers as Tenenbaum and Griffiths in these cases. Moreover, Tenenbaum and Griffiths discuss the choice of a probability density function for the *identity* of the consequential region; so they are not discussing exactly the same question which was taken up by Shepard. (It is curious to us that Tenenbaum and Griffiths use Erlang priors in their examples: after all, in their view, Shepard's 'rational justification' for this choice of prior *relies upon* the assumption of weak sampling. From their point of view, it requires work to show that these priors admit of 'rational justification' under the assumption of strong sampling; and we would add that there are mathematically convenient priors which do not admit of rational justification (see Wallace and Dowe (1999b:334–5)). On an unrelated point, it is also curious to us that Tenenbaum and Griffiths claim that *generality* is more primitive than *similarity* (Section 4.1)—but then go on to say that judgments about similarity are required in order to provide reasonable constraints on generalisation (Section 5). We doubt that there is any neat separation of problems to be made here: see Wallace and Boulton (1973) for an example of the simultaneous analysis of both problems in a Bayesian machine–learning context.)

Perhaps Tenenbaum and Griffiths might reply that they have given an extension of Shepard's work to an area which Shepard himself had not considered: given cases in which the natural assumption is strong sampling, what kinds of probability distributions would we expect creatures to have? However, as Tenenbaum and Griffiths note, there are many other kinds of cases in which various other natural assumptions may or may not be made—and, in the end—though Tenenbaum and Griffiths don't put it this way—perhaps all that they commit themselves to is the claim that the most natural *general* extension of Shepard's work supports nothing more than the hypothesis that Bayesian inference is universal. Yet, if that is right, then it is hard to see that we are getting 'universal, invariant general principles' which reflect the internalisation of 'pervasive and enduring facts about the world'. Perhaps we might hold that we are getting 'universal, invariant, general principles': though even that seems a bit of a stretch. After all, it is one question what are the optimal inferences to be made from given data; it is quite another question how close we should expect evolved creatures to come to these optimal inferences in particular cases. There seems to be little reason to suppose that evolved creatures will be perfect Bayesians across the board; indeed, we know from countless experiments on people that we are very far from being perfect Bayesian reasoners ourselves. Unless we are prepared to hold that you don't count as a 'perceptually advanced mobile organism' unless you are a reliable Bayesian information processor, we see little reason to suppose that there is a *universal* law of the kind which Shepard proposes (even allowing a restriction to cases in which weak sampling can be presupposed). Opponents of Bayesian theories of inference will no doubt also have many reasons for wishing to raise objections here.

Tenenbaum and Griffiths appear to claim that their theory 'uniquely' extends Shepard's work in a Bayesian framework (3.3). However, one can use a Bayesian approach to infer the consequential region itself: see, for example, Wallace and Dowe (1999b: 332–4) for analysis of a related continuous case in which the size of the consequential region is known.

We agree with the implicit suggestion of Barlow (2001) that MML theory might offer the best Bayesian theory of inference—see Wallace and Boulton (1968), Wallace and Freeman (1987); and see Wallace and Dowe (1999a) (and the rest of this 1999 special issue of the *Computer Journal*) for a discussion of the contrast between MML, MDL, and the work of Solomonoff (1964) and others. To the extent that we expect evolution to optimise, we should expect to find MML inferences in nature. No doubt we often do. MML can be made to fit the empirical data described by Shepard (1987). And it can be made to yield the generalisations offered by Tenenbaum and Griffiths. And it has a high degree of mathematical elegance, etc. But we don't think that it would help the general thesis which Shepard defends to espouse MML—even though the suggestion could be quite congenial to him. That Bayesian MML inference is evolutionarily advantageous is true, if it is true, quite independently of facts about the *kinds* of objects that there are in the world.

References

Anderson, J. (1967) *Principles of Relativity Physics* New York: Academic Press

Barlow, H. (2001) "The Exploitation of Regularities in the Environment by the Brain" *Behavioural and Brain Sciences* 24, 3, **THIS VOLUME**

Shepard, R. (1987) "Toward a Universal Law of Generalisation for Psychological Science" *Science* 237, 1317–23

Shepard, R. (1994) "Perceptual–Cognitive Universals as Reflections of the World" *Psychonomic Bulletin And Review* 1, 2-28; reprinted in *Behavioural and Brain Sciences* 24, 3, **THIS VOLUME**

Solomonoff, R. (1964) "A Formal Theory of Inductive Inference" *Information and Control* 7, 1–22, 224–54

Tenenbaum, J. and Griffiths, T. (2001) "Generalisation, Similarity and Bayesian Inference" *Behavioural and Brain Sciences* 24, 3, **THIS VOLUME**

Wallace, C. and Boulton, D. (1968) "An Information Measure for Classification" *Computing Journal* 11, 185–195

Wallace, C. and Boulton, D. (1973) "An information measure for hierarchic classification" *Computer Journal* 16, 254–61

Wallace, C. and Freeman, P. (1987) "Estimation and Inference by Compact Coding" *Journal of the Royal Statistical Society B*, 49, 3, 223–265

Wallace, C. and Dowe, D. (1999a) "Minimum Message Length and Kolmogorov Complexity" *Computer Journal* 42, 4, 270–83

Wallace, C. and Dowe, D. (1999b) "Refinements of MDL and MML Coding" *Computer Journal* 42, 4, 330–7