Focus Article

Ockham's Razor

Nicole Lazar*

Ockham's Razor is the philosophical principle that entities or assumptions should not be introduced unnecessarily. This is often interpreted as stating that one should seek the simplest explanation for a phenomenon. The purposes of this note are to introduce some of the different formulations of the Razor and the controversy surrounding it, to explore the use of Ockham's Razor as a scientific tool, and to draw connections with statistical inference, both frequentist and Bayesian. The note starts with a brief introduction to the life of William of Ockham, to whom the Razor is traditionally and popularly attributed. There is some controversy regarding the use of Ockham's Razor as an inferential tool; this seems to center on the meaning of 'simple explanation' and the issue is explored. Uses of the Razor in science, in statistical model selection, and in Bayesian inference are examined. *WIREs Comp Stat* 2010 2 243–246 DOI: 10.1002/wics.75

Keywords: Bayesian inference; model selection; parsimony

Ockham's Razor is the principle that explanations for a phenomenon should not be multiplied unnecessarily. It is widely and popularly attributed to the medieval philosopher William of Ockham. Although its origins are in the realm of philosophy, Ockham's Razor has solid grounding and acceptance as part of the scientific method; connects at a basic level to the philosophy and practice of statistics both Bayesian and frequentist; and is even casually invoked in day-to-day conversation to settle disagreements.

William of Ockham (c. 1287 to c. 1348), a leading philosopher of the Middle Ages, was born in Ockham, Surrey, outside of London. He joined the Franciscan order at an early age, and was educated in the order's London school, or convent. William started, but did not finish, his theological training at Oxford, earning him the nickname *Venerabilis Inceptor* ('venerable beginner'). Throughout his life, Ockham had a contentious relationship with the Church, going so far as to accuse Pope John XXII of heresy. In 1328, after declaring that the Pope was a heretic due to his views on the poverty of the apostles, Ockham, along with several other members of the Franciscan order, fled to Bavaria for exile; William was subsequently excommunicated. He died in exile around 1348. Although Ockham is today known mostly for his eponymous Razor, he wrote widely on logic, philosophy, physics, theology, and politics; in short, William of Ockham played a major role in all areas of medieval philosophy.[1]

*Correspondence to: nlazar@stat.uga.edu

Department of Statistics, University of Georgia, Athens, GA 30602, USA

DOI: 10.1002/wics.75

QUESTIONS AND CONTROVERSIES

Ockham's Razor has many formulations in the popular and scientific literatures; a typical one is that 'The simpler of two explanations is to be preferred'. More formally, the Razor is expressed as 'Do not multiply entities beyond what is required'. Following Stigler's Law of Eponymy,[2] Ockham was not the first to state this rule of parsimony; versions of it can be found, for example, in the works of Aristotle, Maimonides, Duns Scotus and Aquinas, among others. Indeed, it seems that the statement *Entia non sunt multiplicanda praeter necessitatem* ('entities should not be multiplied unnecessarily'), often attributed to William of Ockham, appears nowhere in his writings, but can instead be found in works that appeared some 300 years after Ockham's death.[3] Similar expressions of the principle ('plurality is not to be assumed without necessity' and 'what can be done with fewer [assumptions] is done in vain with more') can be traced back to Ockham, so the misattribution is perhaps not deeply troubling. Interesting to note, in light of its pervasiveness in the public consciousness, is that contemporaries do not refer to 'Ockham's Razor;' the name was not coined until 500 years after Ockham's death, by William Hamilton, in 1852.[3]

Although what we call Ockham's Razor has gained popularity and acceptance in its various guises, as a philosophical idea it is not without controversy, historically or in modern times. For instance, Ockham's contemporary Walter Chatton took an opposing view, stating that explanations should be added until a phenomenon or proposition could be affirmed (if three entities are not enough, add a fourth, or fifth, and so on; Ref 4). Wears and Lewis[5] raise potential objection to the principle of parsimony (which they equate with Ockham's Razor) in the context of statistical modeling and inference; a simpler explanation of a phenomenon that ignores important variables may lead to poor predictive power and difficulty in validating models. Part of the difficulty evidently centers on the meaning of 'simple'. The objections of Chatton, as well as of Wears and Lewis, involve an understanding of the Razor as demanding utmost—even excessive—simplicity. These 'anti-Razors', as they are sometimes called, approach the parsimony question from the opposite direction, namely, that an entity should be shown to be lacking in importance before it is removed from the explanation. But this does not contradict Ockham's Razor, which stresses 'unnecessary' multiplication of explanations or assumptions. Clearly, an effect that has important explanatory power should not be ignored. Or, as Einstein put it: 'everything should be made as simple as possible, but not simpler'.

THE INTUITIVE USE OF OCKHAM'S RAZOR

One appeal of Ockham's Razor is that it cuts through layers of assumptions and reveals the core of an argument, thereby enabling the impartial outsider to evaluate the validity of competing theories. As a somewhat trivial example, consider the case of the 'Cottingley Fairies', a series of photographs of fairies taken by two young girls in 1917. These photographs, which purportedly show the girls playing with fairies, were widely believed at the time to be authentic; Sir Arthur Conan Doyle, the author of the Sherlock Holmes mysteries, was one believer. But which is the simpler explanation: that fairies do actually exist but have somehow eluded being caught on film (either prior to 1917 or subsequently)? Or that the photographs were fake? The latter explanation requires fewer assumptions (for most people), and hence the Razor indicates it should be favored. Indeed, of course, the photographs were faked, although this was not definitively proven until many years later. Uses such as deciding on the veracity of the Cottingley Fairies correspond to the layman's understanding of Ockham's Razor as 'the simplest theory is the correct one'. This diverges quite substantially from statements of the Razor found in the philosophical literature and in Ockham's own writing, but is close enough for a working definition.

OCKHAM'S RAZOR AS A TOOL OF SCIENCE

The Razor, with its emphasis on simplicity and elegance of explanation, merges well with the current state of scientific practice. Good theories are those that explain all the known facts in a fashion as uncomplicated as possible. Ockham's Razor may therefore be used by scientists to pare away unnecessary hypotheses and arrive at simpler theories. Jefferys and Berger[6] discuss competing explanations early in the 20th century for the motion of the planet Mercury, which was not fully explained by Newtonian physics. One theory held that another, as yet unobserved planet (tentatively named 'Vulcan') existed between the Sun and Mercury, and it was Vulcan's orbit that was causing the perturbances in Mercury. Another theory was that the effect was caused by rings around the Sun. In still another, the Sun was taken to be oblate. Yet another explanation was that the law of gravity was not quite right (see Ref 6 for more details). All of these theories had unknown parameters (for instance, the amount of deviation from the law of gravity, or the size and location of Vulcan), and over time some of them became more and more implausible (for instance, lack of confirmed sightings of Vulcan made the existence of such a planet more and more unlikely as time went on). In 1915, Einstein's theory of general relativity was able to resolve the discrepancy; even more, the perturbance in the motion of Mercury is a consequence of Einstein's theory, lending this explanation an important element of parsimony. Although Ockham's Razor could not have led a scientist to the theory of general relativity, once that theory was proposed the Razor could (and did) lend it increased credibility and scientific appeal. If one theory can explain multiple scientific phenomena, and even predict others (as general relativity has), it is to be preferred over multiple theories, one for each observed outcome. This is yet another view on Ockham's Razor, one that is still consistent with the previous ideas.

THE LAW OF PARSIMONY IN STATISTICAL MODELING

When building statistical models, the law of parsimony is often invoked, at least implicitly. Parsimonious models, with a comparatively small number of parameters, may be preferred because they are easier to interpret and explain; they are less prone to overfitting the particular sample in hand; and they tend to be more generalizable to other samples taken under similar conditions. Model selection criteria such as Akaike Information Criterion (AIC)[7] and Bayes Information Criterion (BIC),[8] to name just two of the many that have been proposed in the statistics literature, penalize models that are too complex, i.e., have too large number of parameters. Smith and Spiegelhalter[9] argue that Bayes Factors for comparing nested linear models work as a sort of 'automatic' Ockham's Razor, selecting the simpler model 'whenever there is nothing to be lost by so doing' (Ref 9, p. 216). Some procedures for variable selection, such as the LASSO,[10] enforce parsimony by shrinking small parameter estimates to zero, thereby creating so-called 'sparse' models. Informally, these procedures and criteria embody the notion that 'simpler models should be preferred until the data justify more complex models' (Ref 11, p. 349), a version of Ockham's Razor.[11] That is, if two competing models both fit the data equally well (an idea that in and of itself perhaps needs further consideration), we should prefer the simpler one, which typically will have fewer parameters. The objection of Wears and Lewis[5] is also addressed by Balasubramanian's formulation, because the latter would require including any variables of potential explanatory power; a model that excluded important predictor variables presumably would not fit the data as well as one that included those variables. Hence, it is again evident that agreement upon what is meant by 'simple' in any given context is crucial for reconciling apparent discrepancies in the scientific use of the Razor.

OCKHAM'S RAZOR AND BAYESIAN STATISTICS

From early days, apparently dating back to the work of Jeffreys (see, for example, Ref 12), connections have been drawn between Ockham's Razor and the Bayesian approach to statistical analysis. These connections are at a philosophical as well as a computational level, and were elaborated on by Jefferys and Berger,[6] Balasubramanian,[11] and others. The context for the connection is again framed in terms of selecting between competing models or hypotheses for the state of nature, but now one is assumed to be (for all practical purposes) true.

As a first attempt at framing Ockham's Razor from a Bayesian viewpoint, one might take the approach that the simpler model is preferable because it has higher prior probability. While this is appealing intuitively, tries to provide an answer to the question of what is meant by 'simple', and accords well with the way science tends to be practiced (we start with simple hypotheses and build them up as needed when more evidence accumulates), the definition is somewhat circular, in the absence of any objective way of assigning the priors. Jeffreys' argument was based on exactly this consideration of prior probabilities.[6] A different tactic is therefore needed. In particular, a definition of simplicity that does not appeal directly to the prior probabilities attached to the competing theories will resolve the circularity. Jefferys and Berger[6] call hypothesis H_0 'simpler' than the alternative H_1 if it makes sharper predictions about future data observations. Their reasoning is that a more complex model will usually be able to explain a wider variety of possible outcomes, whereas a simpler model is more narrowly focused. Consequently, also, a simpler model is easier to invalidate; all that is required are observed data that contradict the prediction of that model (and there will be many such possibilities). On the other hand, if data are observed that match the prediction of the simpler model, its credibility is greatly enhanced, even if the observed data are consistent with the more complicated model as well. In probability terms, the simpler model puts high probability on a small subset of the space of possible (future) observations; the complex model distributes the probability more evenly. These can be thought of as the prior probablities in a Bayesian formulation. Combined with models for the observed data, posterior probabilities for the two hypotheses can be calculated using Bayes' rule. And, further, as discussed by Jefferys and Berger, the simpler model will often have higher posterior probability, so that a comparison using, say, Bayes Factors, would lead to favoring the simpler explanation. Similar results will hold in many settings even if the two competing hypotheses have the same prior probabilities; that is, even with no *a priori* preference for one explanation over the other, Bayesian considerations will often lead to selection of the more parsimonious one. Following Jefferys and Berger,[6] there are three Bayesian interpretations of Ockham's Razor: (1) as in Jeffreys, to set prior probabilities under the logic that experience has shown that simpler theories are more likely to be correct; (2) from the modeling perspective, whereby simple models (those with fewer parameters) make sharper inferences and therefore will have higher posterior probabilities; (3) from a model selection perspective (Bayesian or frequentist), in which criteria such as AIC and BIC favor simpler models.

CONCLUSION

Although there are many statements that are widely recognized as 'Ockham's Razor', many of which differ in emphasis and intent, the main principle is generally preserved. Namely, one should aim for parsimony or simplicity of explanation, and avoid introducing unnecessary assumptions. In these, or similar, formulations, Ockham's Razor is a central part of modern scientific practice. It is central in statistical practice as well, being an implicit driving force behind much of the theory and philosophy of statistical model selection and having justifications in the very foundations of Bayesian inferential procedures.

REFERENCES

1. Spade PV. William of Ockham. In: Zalta EN (ed.). *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition) 2008, http://plato.stanford.edu/archives/fall2008/entries/ockham.
2. Stigler SM. Stigler's law of eponymy. *Trans New York Acad Sci, Ser 2* 1980, 39:147–158.
3. Thorburn WM. The myth of Occam's razor. *Mind* 1918, 27:345–353.
4. Keele R. Walter Chatton. In: Zalta EN (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition) 2008, http://plato.stanford.edu/archives/fall2008/entries/walter-chatton.
5. Wears RL, Lewis RJ. Statistical models and Occam's razor. *Acad Emerg Med* 1999, 6:93–94.
6. Jefferys WH, Berger JO. Sharpening Ockham's razor on a Bayesian strop. Technical Report # 91-44C. Purdue University, 1991.
7. Akaike H. Information theory and an extension of the maximum likelihood principle. In Petrov BN, Czaki F, eds. *Second International Symposium on Information Theory*. Budapest: Akademiai Kiado; 1973, 267–281.
8. Schwarz G. Estimating the dimension of a model. *Ann Stat* 1978, 6:461–464.
9. Smith AFM, Spiegelhalter DJ. Bayes factors and choice criteria for linear models. *J R Stat Soc., Ser A* 1980, 42:213–220.
10. Tibshirani R. Regression shrinkage and selection via the LASSO. *J R Stat Soc., Ser A* 1996, 58:267–288.
11. Balasubramanian V. Statistical inference, Occam's razor, and statistical mechanics on the space of probability distributions. *Neural Comput* 1997, 9:349–368.
12. Robert CP, Chopin N, Rousseau J. Harold Jeffreys' Theory of Probability revisited. *Stat Sci* 2009, 24:141–172.

