INQUIRY · LOGIC AND REASONING

Occam's Razor and the Principle of Parsimony

Occam's razor advises against multiplying assumptions beyond necessity. This article explains where the principle came from, the different things simplicity can mean, why simpler explanations and models often predict better, and when the razor cuts too deep.

ModQu research article · 2,518 words · 25 references (APA 7th edition)

When two explanations account for the evidence equally well, choose the simpler one. This advice, usually called *Occam's razor* or the *principle of parsimony*, is one of the most widely invoked rules of reasoning. Scientists use it to choose between theories, statisticians to choose between models, doctors to choose between diagnoses and managers to choose between explanations of a business problem. Yet it is rarely examined. What exactly counts as simpler? Why should simpler explanations be more likely to be true? And when does the preference for simplicity lead us astray? This article traces the history of the principle, explains the main justifications offered for it and sets out when it helps inquiry and when it misleads.

Origins

The principle is named after William of Ockham, a fourteenth-century English Franciscan philosopher. The phrase most often attributed to him, "entities must not be multiplied beyond necessity", does not appear in his surviving works. W. M. Thorburn (1918) traced its history and showed that the familiar wording was formulated by later writers. Ockham did, however, use the principle frequently, in forms such as "plurality should not be posited without necessity", mainly to argue against unnecessary metaphysical entities (Spade, 1999).

The idea is older than Ockham and was taken up by many later thinkers. Isaac Newton made it the first of his rules of reasoning in natural philosophy: "No more causes of natural things should be admitted than are both true and sufficient to explain their phenomena" (Newton, 1999). The principle played a role in the acceptance of the Copernican system, in which the sun rather than the earth is at the centre, although, as Thomas Kuhn (1957) showed, Copernicus's original system was not in fact much simpler than Ptolemy's in the number of circles it required; its appeal lay in a more unified and natural explanation of certain features of planetary motion.

What does "simpler" mean?

Elliott Sober (2015) argued that there is not one razor but several, and that "simplicity" means different things in different contexts. Alan Baker (2022) distinguishes two main families. *Ontological parsimony* concerns the number and kinds of things a theory says exist: a theory that explains the evidence with fewer types of entity is more parsimonious. *Syntactic simplicity*, or elegance, concerns the number and complexity of a theory's assumptions, laws or parameters.

In practice, simplicity also has other dimensions. An explanation may be simpler because it invokes fewer causes, because it requires fewer special assumptions to fit the evidence, because it unifies phenomena that would otherwise need separate explanations or because it has fewer adjustable parameters. These do not always point the same way. A theory may posit fewer kinds of entity but require more complex laws, or vice versa. Judging simplicity therefore requires saying which kind is meant.

Why prefer simplicity?

Several justifications have been offered for preferring simpler explanations.

Simplicity as a methodological convenience

One view is that simpler theories are easier to work with, test and communicate, without being more likely to be true. Thomas Kuhn (1977) listed simplicity among the values, alongside accuracy, consistency, scope and fruitfulness, that scientists use to choose between theories, and argued that such values guide but do not determine choices. Karl Popper (1959) argued that simpler theories are preferable because they are more easily falsified: they make bolder claims that rule out more possibilities, and so can be tested more severely.

Simplicity and probability

A second view links simplicity directly to probability. On a Bayesian analysis, a complex theory with many adjustable parameters can fit a wide range of possible observations, so it spreads its probability thinly across them. A simpler theory makes sharper predictions, so if those predictions are borne out, the evidence supports it more strongly. William Jefferys and James Berger (1992) showed how this "Bayesian Ockham's razor" arises automatically from probability theory: other things being equal, the simpler theory receives more support from data that both theories can explain. David MacKay (2003) developed the same idea in the context of model comparison.

Simplicity and prediction

A third and highly practical justification comes from statistics and machine learning. Complex models with many parameters can fit the data they were built on very closely, but this close fit often includes the noise as well as the signal. When such models are used to predict new data, they perform worse than simpler models. This is called *overfitting*. Douglas Hawkins (2004) described how overfitting arises and how it can be detected, by testing models on data they were not built on. Hastie et al. (2009) explain the underlying trade-off between bias and variance: simple models may miss real patterns (high bias), while complex models are sensitive to the particular sample (high variance). Figure 1 illustrates the pattern.

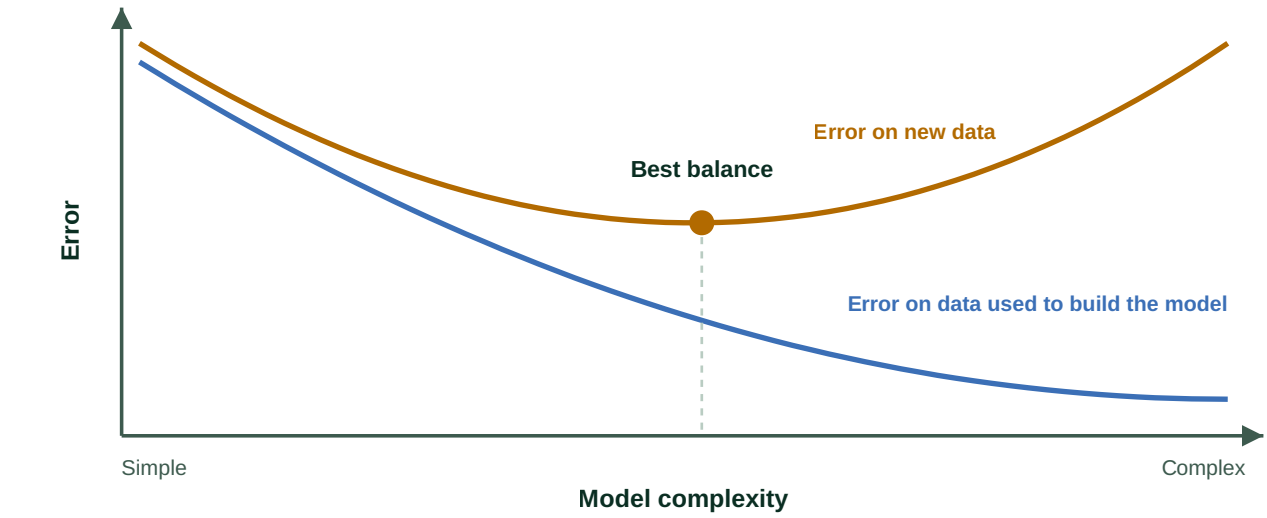


Figure 1. Overfitting: as a model becomes more complex, its fit to the data it was built on keeps improving, but its accuracy on new data eventually gets worse. Schematic; chart by ModQu.

Statisticians have developed formal criteria that build in a penalty for complexity. Hirotugu Akaike (1974) proposed an information criterion that balances how well a model fits against the number of parameters it uses, based on an estimate of how well the model will predict new data. Gideon Schwarz (1978) proposed a related criterion, derived from Bayesian reasoning, that penalises complexity more heavily in large samples. Jorma Rissanen (1978) proposed the *minimum description length* principle: the best model is the one that allows the data to be described most briefly, counting both the description of the model and the description of the data given the model. Forster and Sober (1994) argued that such criteria show how simplicity can be justified as a means to predictive accuracy, rather than as a sign of truth in itself.

Simple models in practice

The evidence that simple models often predict as well as or better than complex ones is striking. Robyn Dawes (1979) showed that simple linear models, in which predictors are given equal weights rather than weights estimated from the data, often predicted outcomes such as students' academic success as well as or better than expert judgement and sophisticated statistical models. Gigerenzer and Brighton (2009) reviewed evidence that simple heuristics, which ignore much of the available information, can outperform complex strategies in uncertain environments, precisely because they are less prone to overfitting.

In forecasting, Makridakis and Hibon (2000) reported the results of a large competition in which many forecasting methods were tested on thousands of time series. Statistically sophisticated methods did not necessarily produce more accurate forecasts than simpler ones. Green and Armstrong (2015) reviewed comparisons of simple and complex forecasting methods and found that complex methods increased forecast error on average, while appearing more credible to users.

When the razor cuts too deep

The principle of parsimony is a guide, not a law, and it can mislead.

The world is not always simple. Simplicity is a reason to prefer one explanation over another only when they explain the evidence equally well. When a simple explanation does not fit the evidence, a more complex one is warranted. Newton's rule includes the word "sufficient" for this reason.

Several causes may operate together. In medicine, Occam's razor suggests seeking a single diagnosis that explains all of a patient's symptoms. A counter-principle, often called *Hickam's dictum*, holds that patients can have as many diseases as they please. Hilliard et al. (2004) described a case in which an apparently unified set of symptoms in fact resulted from several separate conditions, and argued that clinicians, particularly with older patients who often have several conditions at once, must be ready to consider multiple causes. Business problems are similar: a fall in sales may result from several factors at once, and insisting on a single explanation can hide the others.

People overvalue simplicity. Tania Lombrozo (2007) found that people prefer explanations that invoke fewer causes even when the evidence makes a more complex explanation more probable. Lombrozo (2016) reviewed evidence that such explanatory preferences, while often useful, can distort learning. The intuitive appeal of a simple story is not the same as its truth.

Simplicity is not a guarantee. Pedro Domingos (1999) examined the use of Occam's razor in data mining and argued that the belief that simpler models are always more accurate is mistaken. What matters is the ability to generalise to new data, which can be tested directly, and in some settings more complex models generalise better. Kevin Kelly (2007) argued that the best justification for Ockham's razor is not that simpler theories are more likely to be true, but that preferring them leads to the truth with fewer reversals of opinion along the way.

Ad hoc rescues

A particular kind of complexity deserves suspicion: the auxiliary assumption added only to save a theory from contrary evidence. Popper (1959) called such moves *ad hoc*. If a prediction fails and the theory's defenders respond by adding a new assumption that explains away the failure but makes no other testable claim, the theory has become more complex without becoming more informative. Repeated often enough, this makes a theory compatible with any possible observation, and so empty.

Not every added assumption is ad hoc. The difference lies in whether the new assumption can be checked independently. When astronomers noticed irregularities in the orbit of Uranus, they proposed an unseen planet whose gravity would explain them. The assumption added complexity, but it also predicted where the new planet should be, and Neptune was

found close to the predicted position. Forster and Sober (1994) argue that the cost of extra assumptions is best judged by their effect on predictive accuracy: an assumption that improves predictions of new cases earns its place, while one that only accommodates the cases already seen does not.

The same test applies in business analysis. Suppose a model predicts that a price cut will raise volume, and it does not. An analyst can add "customers were distracted by the holiday period" to rescue the model. If that assumption implies something checkable, such as a matching dip in other categories or a recovery once the holiday ends, it can be tested. If it implies nothing else, it is simply a way of avoiding the conclusion that the model was wrong.

Parsimony in reporting

The principle also applies to how findings are communicated. Research reports and dashboards often grow by accretion: each stakeholder asks for another chart, another segment or another metric, until the main finding is buried. A parsimonious report presents the smallest set of results needed to support its conclusions, and places supporting detail where readers can find it if they need it.

This is not a matter of style alone. Every additional comparison in a report is another chance for a pattern that arose by chance to be read as meaningful. The more segments and metrics a report shows, the more likely it is that some will show striking differences that would not recur in a new sample. This is the same problem as overfitting, appearing in the reader's interpretation rather than in a statistical model. Green and Armstrong (2015) noted that complex methods tend to appear more credible to users even when they are less accurate. Detailed reports can create the same false impression of rigour.

Simplicity in a report should not mean hiding complexity that matters. If a result differs sharply between customer groups, the difference is part of the finding and belongs in the summary. The aim is to remove what the conclusions do not depend on, not what complicates them.

Simplicity and understanding

There is one more reason to value simple explanations, separate from truth and prediction. A simple explanation can be understood, remembered and acted on by the people who need it. Lombrozo (2016) notes that explanations do more than describe: they guide what people attend to and how they generalise to new cases. An accurate model that no one in an organisation can explain will rarely change a decision, while a slightly less precise one that managers understand may be used every day.

This creates a real trade-off, and it should be made openly. Where the gain in accuracy from a complex model is small, the simpler model may be the better choice because it will be used and checked. Where the gain is large, the complex model may be worth the effort of explaining it. Kuhn (1977) observed that scientists weigh simplicity against other values, and the same weighing happens in applied research. What matters is that the choice is deliberate, not a default in either direction.

Using the razor well

For anyone weighing explanations or building models, several practical guidelines follow:

1. **Apply it only between explanations that fit the evidence equally well.** Simplicity breaks ties; it does not override evidence.
2. **Say what kind of simplicity is meant.** Fewer causes, fewer assumptions and fewer parameters are different things.
3. **Test predictions on new data.** The best protection against overfitting is to check how well a model or explanation predicts cases it was not built on.
4. **Prefer simple models when data are limited.** With small or noisy data sets, simpler models are less likely to fit noise.
5. **Stay open to multiple causes.** In complex systems, several factors often operate together.
6. **Be suspicious of a story because it is satisfying.** The appeal of a simple explanation is not evidence for it.

A hypothetical example

A retailer sees a sharp fall in online orders over two weeks. The simplest explanation is a technical fault on the website, and an outage is found on one of the days. But the fault lasted only a few hours, while the fall lasted two weeks. Applying the razor properly means recognising that the single cause is not sufficient to explain the evidence. Further investigation finds two additional factors: a competitor's promotion during the same period and a change in a search engine's ranking that reduced traffic. The most parsimonious explanation that fits all the evidence involves three causes, not one. Each requires a different response. A team that had stopped at the outage would have fixed a problem that had already ended and missed the two that were still operating. A test of each explanation against the evidence, including the timing and the channels affected, is what separates a parsimonious account from an incomplete one.

Conclusion

Occam's razor captures a sound intuition: do not add complexity that the evidence does not require. Its justifications are strongest where simplicity improves prediction, as in the choice between statistical models, and where it makes theories easier to test. But simplicity is a guide for choosing between explanations that fit the evidence equally well, not a substitute for evidence. Used carelessly, the razor cuts away real causes and rewards satisfying stories over accurate ones. Used well, it keeps inquiry focused on explanations that are as simple as possible, and no simpler.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Baker, A. (2022). Simplicity. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2022 ed.). Stanford University. https://plato.stanford.edu/archives/sum2022/entries/simplicity/

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*(7), 571–582. https://doi.org/10.1037/0003-066X.34.7.571

Domingos, P. (1999). The role of Occam's razor in knowledge discovery. *Data Mining and Knowledge Discovery, 3*(4), 409–425. https://doi.org/10.1023/A:1009868929893

Forster, M., & Sober, E. (1994). How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *The British Journal for the Philosophy of Science, 45*(1), 1–35. https://doi.org/10.1093/bjps/45.1.1

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science, 1*(1), 107–143. https://doi.org/10.1111/j.1756-8765.2008.01006.x

Green, K. C., & Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research, 68*(8), 1678–1685. https://doi.org/10.1016/j.jbusres.2015.03.026

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences, 44*(1), 1–12. https://doi.org/10.1021/ci0342472

Hilliard, A. A., Weinberger, S. E., Tierney, L. M., Jr., Midthun, D. E., & Saint, S. (2004). Occam's razor versus Saint's triad. *New England Journal of Medicine, 350*(6), 599–603. https://doi.org/10.1056/NEJMcps031794

Jefferys, W. H., & Berger, J. O. (1992). Ockham's razor and Bayesian analysis. *American Scientist, 80*(1), 64–72.

Kelly, K. T. (2007). Ockham's razor, empirical complexity, and truth-finding efficiency. *Theoretical Computer Science, 383*(2–3), 270–289. https://doi.org/10.1016/j.tcs.2007.04.009

Kuhn, T. S. (1957). *The Copernican revolution: Planetary astronomy in the development of Western thought*. Harvard University Press.

Kuhn, T. S. (1977). Objectivity, value judgment, and theory choice. In *The essential tension: Selected studies in scientific tradition and change* (pp. 320–339). University of Chicago Press.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology, 55*(3), 232–257. https://doi.org/10.1016/j.cogpsych.2006.09.006

Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences, 20*(10), 748–759. https://doi.org/10.1016/j.tics.2016.08.001

MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge University Press.

Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results, conclusions and implications. *International Journal of Forecasting, 16*(4), 451–476. https://doi.org/10.1016/S0169-2070(00)00057-1

Newton, I. (1999). *The Principia: Mathematical principles of natural philosophy* (I. B. Cohen & A. Whitman, Trans.). University of California Press. (Original work published 1687)

Popper, K. R. (1959). *The logic of scientific discovery*. Hutchinson.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica, 14*(5), 465–471. https://doi.org/10.1016/0005-1098(78)90005-5

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Sober, E. (2015). *Ockham's razors: A user's manual*. Cambridge University Press. https://doi.org/10.1017/CBO9781107705937

Spade, P. V. (Ed.). (1999). *The Cambridge companion to Ockham*. Cambridge University Press. https://doi.org/10.1017/CCOL052158244X

Thorburn, W. M. (1918). The myth of Occam's razor. *Mind, 27*(107), 345–353. https://doi.org/10.1093/mind/XXVII.3.345