Model complexity as the mean distance between a model and arbitrary data

John C. Dunn

University of Western Australia

May, 1999

Address for Correspondence:

John C. Dunn
Department of Psychiatry & Behavioural Science
University of Western Australia
Nedlands WA 6907
AUSTRALIA

Tel: (+618) 9346 2251
Fax: (+618) 9346 3828
email: jdunn@cyllene.uwa.edu.au
Homepage: http://www.general.uwa.edu.au/u/kraepeln/jdunn/jdunn.htm

Abstract

Model complexity is conceptualised as the capacity of a model to fit any conceivable data set. A model can be represented geometrically as a low dimensional “response surface” embedded in a higher dimensional outcome space in which data are represented as a point or set of points. The fit of the model to a data point is given by the minimum distance between the point and the response surface. Model complexity can thus be thought of as the extent to which the response surface is “close to” arbitrary points in outcome space. If the extension of outcome space can be assumed to be bounded, complexity can be operationalised as the mean minimum distance, defined as the average minimum squared distance between an arbitrary data point in outcome space and the model’s response surface. It may also be expressed as a dimensionless quantity called the scaled mean minimum distance. For linear models, theoretical values for the scaled mean minimum distance and the variance of the scaled minimum distance can be readily obtained and compared against empirical estimates obtained from fits to random data. The approach is applied to resolving the question of the relative complexity of the linear integration model (LIM) and the fuzzy logic of perception model (FLMP), both of which have been the subject of controversy in the field of depth perception. It is concluded that the two models are equally complex.

The aim of this paper is to explore a heuristic approach to assessing model complexity in terms of the capacity of a model to fit arbitrary patterns of data. Consider two models, *A* and *B*, that purport to account for the same set of data, *D*. If the models contain free parameters, then these are usually adjusted in order to minimise the discrepancy between the predictions of each model and the data. The fit of a model is then indexed by the size of this discrepancy which, if large enough, may lead to its rejection. However, suppose models *A* and *B* fit the data equally well. Does this mean that both models are equally good accounts of the data? Only if they have an equal propensity to fit any conceivable set of data. In other words, only if they are equally complex. The idea is that models differ in their propensity to fit any conceivable data. Some models, by their nature, will never provide a poor fit to any data. In the extreme, a model may be capable of fitting *exactly* every conceivable outcome in which case it is generally regarded as vacuous, uninformative, or unscientific (Popper, 1959). Such models are never preferred, despite the fact that they fit the data perfectly. Thus model selection is a function of the actual fit of a model to data and its prior propensity to fit any conceivable data.

Following Bamber and van Santen (1985), a model, *M*, can be defined as an ordered triple, (P, F, Q), where, for positive integers *m* and *n*, $P \subseteq R^m$ is a *parameter domain* consisting of all conceivable combinations of *m* parameter values of the model, *F* is a *prediction function* defined on *P* such that $F(P) \subseteq Q$, and $Q \subseteq R^n$ is an *outcome space* consisting of all conceivable combinations of the values of *n* different *outcome quantities*. The set $R = F(P)$ is called the model's *prediction range* and defines an *m*-dimensional "response surface" in *n*-dimensional outcome space. For an appropriate choice of metric defined on *Q*, it is possible to associate with each point $q \in Q$, a positive number, $d_M(q)$, defined as the minimum distance between *q* and *R*. Let D_M be the set of minimum distances for each point in *Q* relative to the prediction range of *M*. When an appropriate experiment to test *M* is conducted, a set of outcomes, $Y \subset Q$, is defined. For each $y \in Y$, there is a corresponding minimum distance, $d_M(y) \in D_M$, which constitute the set $D_M(Y) \subset D_M$. The sizes of the elements of this set reflect the ability of the model to account for the data.

Consider two models, *A* and *B*, represented as two low-dimensional response surfaces in the same outcome space, *Q*. Associated with each model is its set of minimum distances, D_A and D_B, corresponding to the smallest distances between each point in *Q* and the response surfaces of *A* and *B*, respectively. Following an experiment, the set of observations, *Y*, selects two corresponding subsets of minimum distances, $D_A(Y) \subset D_A$ and $D_B(Y) \subset D_B$. If the elements of $D_A(Y)$ are generally less than the corresponding elements of $D_B(Y)$, then model *A* provides a better fit to the data than model *B*. Often it is further concluded that this fact counts as evidence *in favour* of model *A* which is then preferred over model *B* as an account of the data. However, this conclusion only follows if the distribution of distances in D_A is the same as the distribution of distances in D_B. If the distributions of minimum distances associated with the two models are different, then a simple choice of the "better" model becomes problematic. Suppose that all the elements of D_A are less than the corresponding elements of D_B, then the observation that the elements of $D_A(Y)$ are also small relative to the corresponding elements of $D_B(Y)$ no longer counts as compelling evidence that *A* is the better model. This observation would be made no matter what the outcome of the experiment. This suggests that the relationship of the response surfaces of each model to the set of all conceivable experimental outcomes is critical to model evaluation.

Myung and Pitt (1997) define model complexity as "the flexibility inherent in a model that enables it to fit diverse patterns of data" (p.80). The present analysis suggests that

complexity, in this sense, be identified with the distribution of minimum distances associated with a model. A useful measure to characterise the relevant aspect of this distribution is the mean. The proposal is therefore that the complexity of a model can be usefully indexed by the mean of its set of minimum distances. This idea is illustrated in Figure 1.

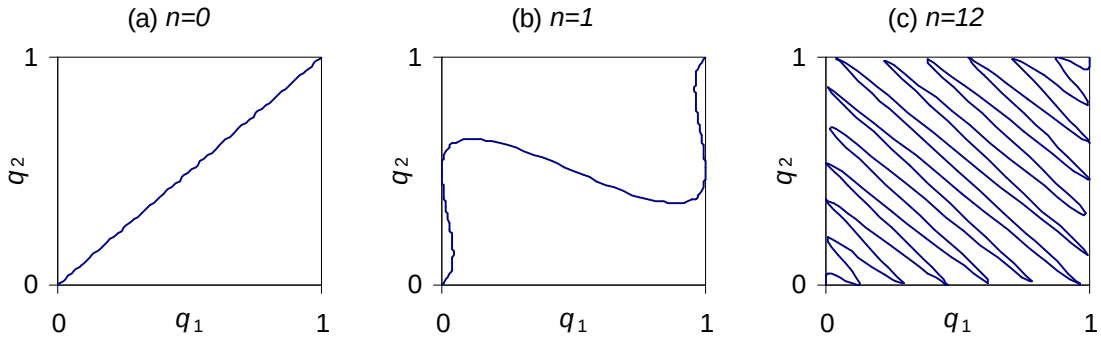

Figure 1: Response surfaces of three models generated from equation 1.

Figure 1 illustrates the one-dimensional response surfaces of three models, each embedded in a two-dimensional outcome space. The prediction function of each model has been generated from the following equation,

$$\begin{aligned} q_1 &= t - \sin\left(2\pi n t + \frac{1}{2} - \left|t - \frac{1}{2}\right|\right) \\ q_2 &= t + \sin\left(2\pi n t + \frac{1}{2} - \left|t - \frac{1}{2}\right|\right) \end{aligned} \tag{1}$$

for $0 \leq t \leq 1$, and for three different values of n, specifically, $n = 0$, $n = 1$, and $n = 12$.

Equation 1 is not meant to signify a meaningful model of anything, and serves only to illustrate the present approach to model complexity. As is apparent from Figure 1, the response surfaces of the three models lie "close to" increasingly greater areas of outcome space. In the limit, as $n \to \infty$, the entire space would be filled. As is also apparent, the mean minimum distance between any arbitrary point and the response surface decreases as n increases. In other words, model complexity, or the propensity of the model to fit any arbitrary pattern of data, *increases* as n increases. In the limit, the mean minimum distance is zero and the corresponding model is infinitely complex.

A more complete picture of the relationship between the mean minimum distance and n for the set of models is given in Figure 2. As it turns out, the relationship between mean minimum distance and n is non-monotonic although mean minimum distance declines rapidly for values of n greater than about 0.5. Thus, the different models differ in their complexity in the sense that they differ in their mean minimum distance to arbitrary data. This complicates model selection. If a less complex model provides at least as good a fit to data as a more complex model, it can be preferred. However, if the more complex model provides a better fit, it is not clear that it should be preferred.

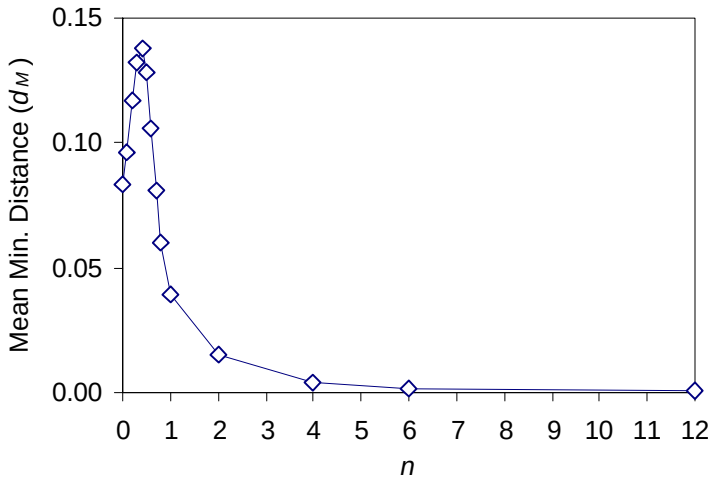

Figure 2: Mean minimum distance as a function of *n* in equation 1.

Myung and Pitt (1997) have suggested that three aspects of a model affect its complexity. These are, the form of the prediction function, *F*, the number of free parameters, *m*, and the extension of the parameter domain, *P*. Each of these aspects also affects the mean minimum distance of a model. First, as Figure 2 illustrates, the form of the prediction function may have a considerable effect on mean minimum distance for constant *m* and *P*. Each of the models examined has the same number of free parameters (one) and the same domain, *P* = [0,1], yet their mean minimum distances vary considerably. Second, as the number of free parameters of a model increases, all things being equal, the model will extend further into outcome space and hence be "closer" to more points in that space. In the limit, if the number of independent parameters equals the number of observations, the response surface will occupy the entire outcome space and the mean minimum distance will be zero. Third, if the parameter domain is restricted in some way, not all of the response surface may be realised. This will also lead to an overall increase in mean minimum distance.

The scaled mean minimum distance

In this section, the idea of model complexity as mean minimum distance is examined more formally and a normalised version, called the *scaled mean minimum distance*, is derived. In the approach to be outlined, the sum of squared difference is used as the distance metric although other functions are possible. Let *M* be a model and let $q \in Q$ be a point in outcome space and let $r \in R$ be a point in the prediction range (or response surface). Then the sum of squared difference between *q* and *r* is,

$$SSD(q, r) = \sum_{i}^{n} (q_i - r_i)^2 \qquad (2)$$

The goodness-of-fit of *M* is evaluated by finding a point, $\hat{r} \in R$, such that equation 2 is a minimum. This is the *minimum distance of M at q*,

$$d_M(q) = \min_{r \in R} SSD(q, r) = SSD(q, \hat{r}) \tag{3}$$

An obvious indicator of the overall "closeness" of the prediction range of a model, *M*, to arbitrary points in the outcome space is the *mean minimum distance of M*, d_M, defined as,

$$d_M = \frac{\int \cdots \int d_M(q)\, dq_1 dq_2 \ldots dq_n}{\int \cdots \int dq_1 dq_2 \ldots dq_n} \tag{4}$$

In order to evaluate equation 4, it is necessary to place bounds on the range of possible outcome values. That is, each outcome quantity, q_i, is deemed to lie between a lower bound, a_i, and an upper bound, b_i. If all of the quantities are commensurable or correspond to measurements of a single variable under *n* different conditions, then it is often possible to specify a common upper and lower bound. Let *a* be the common lower bound and let *b* be the common upper bound. The denominator of equation 4 then simplifies to $(b - a)^n$, and,

$$d_M = \frac{1}{(b-a)^n} \int_a^b \int_a^b \cdots \int_a^b d_M(q)\, dq_1 dq_2 \ldots dq_n \tag{5}$$

As defined by equation 5, the size of d_M depends upon the choice of a measurement scale. Since it is convenient to think of complexity as scale independent, it is useful to express d_M as a dimensionless quantity by re-scaling the outcome space to the unit (hyper)cube, effectively setting $a = 0$ and $b = 1$. This is called the *scaled mean minimum distance of M*, d_M, defined as,

$$d_M = \frac{d_M}{(b-a)^2} \tag{6}$$

In some cases it may be possible to obtain d_M directly from equation 3 by integrating each component of *q* over the interval [0,1]. That is,

$$d_M = \int_0^1 \cdots \int_0^1 d_M(q)\, dq_1 \ldots dq_n \tag{7}$$

Complexity of a linear model

A linear model is of the form, $M = (P, F, Q)$, where $P \subseteq R^m$, $Q \subseteq R^n$, and F: $R^m \rightarrow R^n$ is a linear transformation. Let **p** be a column vector of length *m* corresponding to the point $p \in P$. Similarly, let **r** be a column vector of length *n* corresponding to the point $r \in R = F(P)$, the prediction range. Then, $\mathbf{r} = \mathbf{Ap}$, where **A** is an $n \times m$ matrix of *F*. Linear models are simple, tractable and enjoy some popularity among model builders. Since they are often the least complex of models, they can serve as a baseline against which the complexity of other models may be compared (although see Figure 2 for an exception). For this reason, equations

for the mean and variance of the scaled minimum distance of an arbitrary linear model are derived.

Scaled mean minimum distance

Let $M = (P, F, Q)$ be a linear model such that $P = R^m$, $Q = [0,1]^n$, F: $R^m \rightarrow R^n$ is a linear transformation and $\mathbf{A}$ is a matrix of F. Let $\mathbf{q}$ be a column vector of length n corresponding to the point $q \in Q$ and let $\hat{\mathbf{r}}$ be a column vector of length n corresponding to the point $\hat{r} \in R$ that minimises the sum of squared difference, $SSD(q, r)$. Then by elementary matrix algebra,

$$\hat{\mathbf{r}} = \mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'\mathbf{q}$$

where $\mathbf{A}'$ is the transpose of $\mathbf{A}$. The minimum distance of M at q is given by,

$$d_M(q) = \mathbf{q}'\mathbf{B}\mathbf{q}$$

where, $\mathbf{B}$ is the orthogonal projection matrix,

$$\mathbf{B} = \mathbf{I} - \mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'$$

From these equations, the scaled mean minimum distance of M, d_M, can be calculated. From equation 7,

$$\begin{aligned} d_M &= \int_0^1 \cdots \int_0^1 d_M(q)\, dq_1 \ldots dq_n \\ &= \int_0^1 \cdots \int_0^1 \mathbf{q}'\mathbf{B}\mathbf{q}\, dq_1 \ldots dq_n \\ &= \int_0^1 \cdots \int_0^1 \left[\sum_i^n \sum_j^n q_i q_j b_{ij}\right] dq_1 \ldots dq_n \end{aligned}$$

Rearranging the terms inside the square brackets,

$$\sum_i^n \sum_j^n q_i q_j b_{ij} = \sum_i^n q_i^2 b_{ii} + \sum_i^n \sum_{j \neq i}^n q_i q_j b_{ij}$$

and integrating with respect to each component of q yields,

$$\begin{aligned} d_M &= \tfrac{1}{3}\sum_i^n q_i^3 b_{ii} \prod_{k \neq i}^n q_k + \tfrac{1}{4}\sum_i^n \sum_{j \neq i}^n q_i^2 q_j^2 b_{ij} \prod_{k \neq i,j}^n q_k \Big|_0^1 \\ &= \tfrac{1}{3}\sum_i^n b_{ii} + \tfrac{1}{4}\sum_i^n \sum_{j \neq i}^n b_{ij} \\ &= \tfrac{1}{12}\,\mathrm{tr}(\mathbf{B}) + \tfrac{1}{4}\,\Sigma(\mathbf{B}) \end{aligned} \qquad (8)$$

where, $\mathrm{tr}(\mathbf{B})$ is the trace of $\mathbf{B}$ and $\Sigma(\mathbf{B})$ is the sum of the elements of $\mathbf{B}$.

As an example, consider the following linear model,

$$\begin{aligned} r_1 &= p_1 \\ r_2 &= p_1 + p_2 \\ r_3 &= p_1 + p_3 \\ r_4 &= p_1 + p_2 + p_3 \end{aligned} \tag{9}$$

That is, $\mathbf{r} = [r_1, r_2, r_3, r_4]'$, $\mathbf{p} = [p_1, p_2, p_3]'$, and the design matrix is,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

This yields the following orthogonal projection matrix,

$$\mathbf{B} = \frac{1}{4}\begin{bmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

Applying equation 8, we find that $d_M = 1/12$. In fact if $\mathbf{A}$ defines a response surface in $Q = [0,1]^n$ that includes the vector, $\mathbf{1} = [1,1,1\ldots]'$, then equation 8 simplifies further. In this case,

$$d_M = \frac{n-m}{12}$$

Variance of the scaled minimum distances

For the model M, defined above, let D_M be the set of scaled minimum distances associated with each point, $q \in Q$. Let V_M be the variance of the elements of D_M. This is called the variance of the scaled minimum distances of M, and is defined as,

$$V_M = \int_0^1 \cdots \int_0^1 \left[d_M(q) \right]^2 dq_1 \ldots dq_n - d_M^2 \tag{10}$$

The first term on the right-hand side of equation 10 can be obtained using the same logic as that used to obtain the mean, d_M, using the orthogonal projection matrix, $\mathbf{B}$. That is,

$$d_M(q) = \mathbf{q}'\mathbf{B}\mathbf{q} = \sum_i \sum_j q_i q_j b_{ij}$$

In squaring $d_M(q)$, each term on the right hand side of this equation is multiplied by every other term and summed. That is,

$$[d_M(q)]^2 = \sum_i \sum_j \sum_k \sum_l q_i q_j q_k q_l b_{ij} b_{kl}$$

and integrating with respect to each component of q yields,

$$\int_0^1 \cdots \int_0^1 [d_R(q)]^2 \, dq_1 \ldots dq_n = \sum_i \sum_j \sum_k \sum_l c_{ijkl} b_{ij} b_{kl} \qquad (11)$$

Where the coefficient, c_{ijkl}, varies depending upon the relationship of the numerical value of the subscripts, i, j, k, and l. Specifically, if all the subscripts are different then $c_{ijkl} = 1/16$, if exactly two subscripts are equal then $c_{ijkl} = 1/12$, if one pair of subscripts are equal and the remaining pair are equal to each other but different from the first pair then $c_{ijkl} = 1/9$, if exactly three subscripts are equal then $c_{ijkl} = 1/8$, and if all four subscripts are equal then $c_{ijkl} = 1/5$.

On substituting equations 8 and 11 into equation 10, the variance of the scaled minimum distance is obtained. Applying this formula to the model specified by equation 9 yields the value, $V_M \approx 0.012$.

Complexity of the Linear Integration Model and the Fuzzy Logical Model of Perception

The issue of the relative complexity of different models was highlighted in a recent debate concerning two models of depth perception. The two models specify how depth cues are integrated to form a judgment of overall depth. According to the *linear integration model* (LIM) proposed by Cutting and his co-workers (Bruno & Cutting, 1988; Cutting, Bruno, Brady & Moore, 1992), parameters representing the value of different depth cues are combined additively. According to the *fuzzy logical model of perception* (FLMP) proposed by Massaro and his co-workers (eg., Massaro, 1998; Massaro & Cohen, 1993), a corresponding set of parameters are combined non-linearly using equations derived from the theory of fuzzy logic (see Appendix). Cutting *et al*. (1992), investigated how well the two models fit data collected by Bruno and Cutting (1988) consisting of judgments of overall depth obtained under 16 different combinations of depth cues. Overall, the two models were found to fit these data equally well.

On the face of it, both the LIM and the FLMP appear to provide adequate explanations of the experimental data. However, Cutting *et al* suggested that the relatively good fit of the FLMP was a consequence of its greater propensity to fit any conceivable set of data. If the FLMP has greater complexity, in this sense, then a good fit to the data may not be unexpected and the fact that it provides as good a fit to the data as a less complex model, should be counted as evidence *in favour* of the simpler model.

The LIM is a linear model defined on the outcome space, $Q = [0,1]^n$. Consequently, the mean and variance of the set of scaled minimum distances can be calculated. In the version of the model analysed by Cutting *et al*. (1992), subjects were presented with stimuli under 16 different experimental conditions corresponding to the orthogonal combination of the presence and absence of four depth cues. The cues were relative size, height in the visual field, occlusion, and motion perspective. In addition, it was assumed that a depth cue due to background factors was present in every condition. Thus, the LIM can be defined as a linear model that maps points in a 5-dimensional parameter domain, $P = R^5$, onto points in a 16-dimensional prediction range, $R \subset Q$, where $Q = [0,1]^{16}$. Using equations 8 and 10, the mean and variance of the minimum distances can be obtained. These are,

$$d_{LIM} \approx 0.927$$
$$V_{LIM} \approx 0.090$$

In contrast to the LIM, the FLMP is a non-linear model. This means that the orientation of its response surface relative to a basis of the outcome space varies from point to point. It is because this orientation is constant for a linear model that the orthogonal projection matrix can be used to calculate the minimum distance of a point to the surface. For non-linear models, there is often no analytic solution to the minimum distance of a point which must then be estimated using search algorithms.

Estimates of the mean minimum distance of the two models were obtained by Cutting *et al.* (1992) by calculating the minimum sum of squared difference for each of 1,000 arbitrary data points generated over the interval [0.001, 0.999]. The estimates of d_{LIM} and d_{FLMP} were 0.932 and 0.924, respectively. Using equation 6, these can be converted to estimates of the scaled mean minimum distances, d_{LIM} and d_{FLMP}, yielding values of 0.936 and 0.928, respectively. The first of these is significantly larger than the theoretical value, $z = 2.01$, $p = 0.044$, while the second is not significantly different from the theoretical value for the LIM, $z = 1.17$, $p = 0.244$. This leads to slightly odd conclusion that the FLMP is apparently more complex than the LIM, but only because the observed scaled mean minimum distance for the LIM is larger than expected. Since the LIM cannot, in principle, differ from its theoretical value, the safest conclusion is that there is no evidence of a difference in complexity between the LIM and the FLMP based upon their scaled mean minimum distances.

Although the scaled mean minimum distances of both the LIM and the FLMP are close to the theoretical value for a linear model, Cutting *et al* observed that the FLMP provided a better fit for 608 of the random data points. This is a highly significant departure from the expected number of 500. Similarly, the small difference in the mean minimum distance between the models was also statistically significant using a related-samples *t*-test. These observations suggest that the FLMP does enjoy a slight advantage over the LIM in terms of its propensity to fit arbitrary data. This could occur if the distributions of minimum distances for the two models are different. Specifically, this pattern would result if, for most points in the outcome space, the minimum distance of the FLMP is slightly less than the LIM, while for a small proportion of points, it is considerably longer. In order to investigate this possibility, Dunn (in press) investigated the distribution of minimum distances for the two models by calculating their cumulative proportion ogives. A sample of 500 random data points was generated and fitted by the LIM and the FLMP. The results found by Cutting *et al* were essentially replicated. The values of the mean minimum distances were very similar to those

obtained by Cutting *et al*, the FLMP fit the data better than the LIM on 294 occasions, a similar proportion to that found by Cutting *et al*. (0.59 compared to 0.61), and the small difference between the mean minimum distances for the two models was also highly significant.

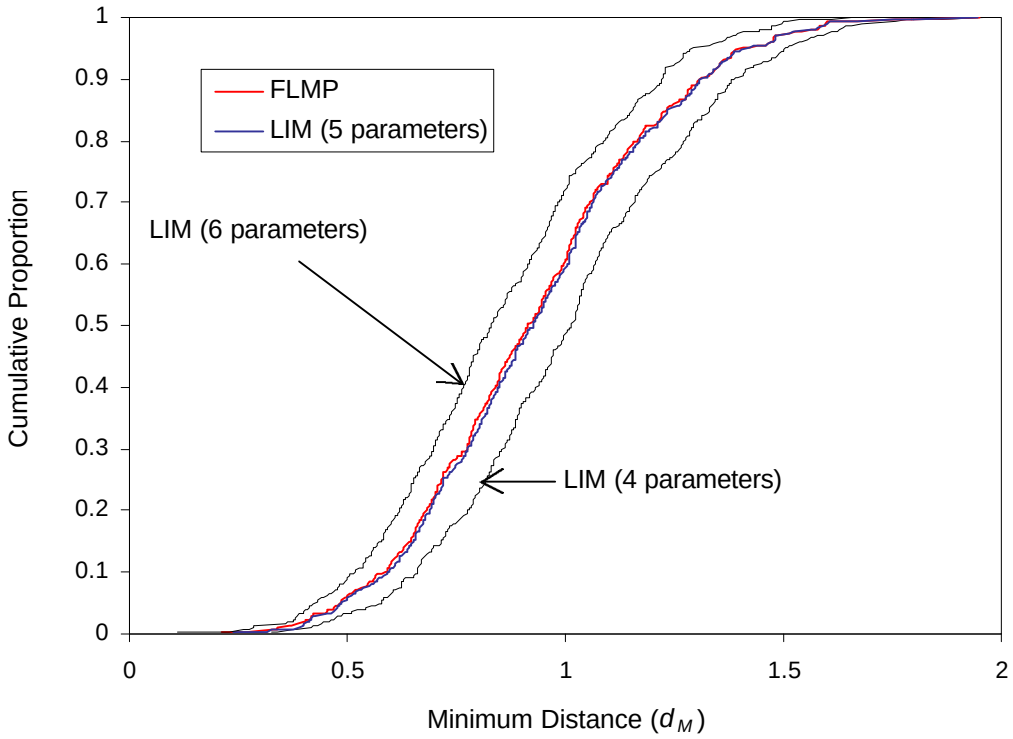

Figure 3: Cumulative proportion ogives as a function of minimum distance for the FLMP, the LIM, and alternative 6-parameter and 4-parameter versions of the LIM (from Dunn, in press).

Figure 3 presents the cumulative proportion ogives for the LIM and the FLMP found by Dunn (in press). If the FLMP provides a generally better fit to random data, its ogive should be shifted leftwards in Figure 3 relative to the LIM. It is apparent that while there is some evidence of such a shift, by and large the two functions are almost coincident. In order to appreciate the magnitude of this small difference, Dunn (in press) plotted the cumulative proportion ogives for a four-parameter and six-parameter version of the LIM. In the four-parameter version, two of the four parameters associated with variable depth cues were constrained to be equal, effectively eliminating one parameter from the model and shifting its ogive substantially to the right relative to the five-parameter LIM. In the six-parameter version, an additional parameter, corresponding to the interaction of the five free parameters, was added, resulting in a substantial shift of the ogive to the left. In comparison to these changes, any difference in complexity between the standard five-parameter LIM and the FLMP is negligible.

Conclusions

The aim of the present article has been to explore a definition of model complexity in terms of the prior propensity of a model to fit arbitrary data sets. Used in this sense, complexity is similar to the idea of "fittability", or the extent to which a model is already "close" to any conceivable data, prior to any data being actually collected. In this sense, complexity is similar to the idea of *logical probability* (Popper, 1959). On this view, if theory *A* is inconsistent with *X*% of all possible outcomes and theory *B* is inconsistent with *Y*% of all possible outcomes, and if $X > Y$, then theory *A* has a lower logical probability than theory *B*. According to Popper's falsificationist approach, theories with low logical probability are more falsifiable and hence constitute better scientific theories. The problem with this idea is that all quantitative theories are inconsistent with an infinite number of outcomes. As Bamber and Van Santen (1985) have shown, unless a quantitative model is saturated (i.e. is inconsistent with no conceivable outcome), the ratio of the set of consistent outcomes to the set of inconsistent outcomes is always zero. In contrast, the concept of model complexity, or "fittability", overcomes this difficulty. While the prior probability that a randomly selected data point lies in the prediction range of a quantitative model is always zero, the minimum distances between points in the outcome space and the prediction range vary continuously. If the mean distance between a random data point and a model's prediction range is greater for model *A* than for model *B*, then model *A* is less complex and, by implication, more falsifiable.

A crucial limitation of the present approach is that outcome space is uniformly bounded. Only by scaling the lower and upper bounds to the unit hypercube, is it possible to define the statistics, d_M and V_M. A linear model is always scalable in this way, but other non-linear models may not be. Fortunately, the FLMP is specifically defined on the interval [0,1] and so the mean and variance of the scaled minimum distances are meaningful. An additional related assumption is that arbitrary data points are uniformly distributed within the bounded outcome space. It is conceivable that both of these assumptions may be relaxed. That is, within an unbounded outcome space, *Q*, we define a prior distribution of points, $H(Q)$, that can take any form. The complexity of a model is then related to the normalised mean minimum distance between this set of points and the model's prediction range.

Pursuing this idea, there are two ways in which $H(Q)$ may be defined. The first is that the structure of the measurement system may impose logical limits on which points in *Q* are feasible. For example, for the two-dimensional space portrayed in Figure 1, it may be the case that q_2 is necessarily always greater than q_1, simply by virtue of how these quantities are defined. Consequently, the set of feasible points, $H(Q)$, is limited to the upper triangular region of *Q* bounded by the line $q_1 = q_2$. The second way in which $H(Q)$ may be defined is more problematic. In this case, the set of conceivable outcomes is determined by a prior probability distribution. That is, some regions of *Q* are defined as more likely than others with the result that the minimum distances between points in these regions and the model's response surface will tend to dominate calculation of d_M. While this approach is feasible, it raises a logical problem due to the fact that expectations concerning the distribution of possible data are not themselves atheoretical. By assuming a uniform distribution within a bounded outcome space, model complexity can be evaluated independently of any actual outcome. That is, a complex model that is "close" to any conceivable outcome can be said to enjoy, *a priori*, an advantage over a less complex model that is, on average, further from these data. If model complexity is evaluated in the context of prior constraints on the distribution of possible outcomes, then model complexity and model fit become inextricably linked. In the limit, as the set of probable outcomes is constrained to approximate more and

more the way the world actually is, model complexity will cease to have any meaning. A model will be highly "fittable" and thus more complex simply because it is the correct model.

It should also be noted that the present analysis is dependent upon the choice of measurement scale adopted. Let $T : [0,1]^n \rightarrow [0,1]^n$ be a transformation of scaled outcome space that maps Q onto $T(Q)$. If model A is more complex than model B in Q, there is no guarantee that it will also be more complex in $T(Q)$. For example, the FLMP is slightly more complex than the LIM over much, if not all, of Q due to non-linearities in its response surface. If the pseudo-probability measures of Q are converted to logits then the FLMP is a linear model in transformed space and the LIM is a non-linear model (see Appendix). Although the logit transformation, $L : [0,1]^n \rightarrow R^n$, does not allow calculation of the mean scaled minimum distance over the entire outcome space, $L(Q)$, it may be expected that the relative complexities of the two models in a bounded region of $L(Q)$ may well be reversed. This implies that complexity may not be an inherent characteristic of a model, but also depends on geometric properties of the chosen outcome space.

Finally, the present approach is designed primarily as a heuristic for thinking about model complexity, model fit, and model preference. The relationship between these concepts is illustrated in Table 1. If model A and model B are equally complex then, all else being equal, preference is determined by fit. In the case of comparison between the LIM and FLMP, they are equally complex and appear to fit the selected data equally well. Therefore, at least in relation to the chosen data set, neither is to be preferred over the other. If model A is more complex than model B, then this rule no longer applies. If the two models fit the data equally well then the less complex model should be preferred. The major problem, not addressed by the present approach is the case in which the more complex model fits the data better. In this situation, caution should be exercised in reaching a judgment of preference and sophisticated model selection procedures, such as the one proposed by Myung and Pitt (1997), explored.

Table 1: The relationship between model complexity, model fit, and model preference.

Complexity[1]	Fit[2]	Preference
$A = B$	$A < B$	B
	$A = B$	Neither
	$A > B$	A
$A > B$	$A < B$	B
	$A = B$	B
	$A > B$	Unknown

1. $A > B$ means that A is more complex than B.
2. $A > B$ means that A fits the data better than B.

References

Bamber, D. & van Santen, J.P.H. (1985). How many parameters can a model have and still be testable? *Journal of Mathematical Psychology, 29*, 443-473.

Bruno, N. & Cutting, J.E. (1988). Minimodularity and the perception of layout. *Journal of Experimental Psychology: General, 117*, 161-170.

Cutting, J.E., Bruno, N., Brady, N.P., & Moore, C. (1992). Selectivity, scope, and simplicity of models: A lesson from fitting judgments of perceived depth. *Journal of Experimental Psychology: General, 121*, 364-381.

Dunn, J.C. (in press). Model complexity: The fit to random data reconsidered. *Psychological Research*.

Massaro, D.W. & Cohen, M.M. (1993). The paradigm and the fuzzy logical model of perception are alive and well. *Journal of Experimental Psychology: General, 122*, 115-124.

Massaro, D.W. (1998). *Perceiving talking faces*. Cambridge MA: MIT Press.

Myung, I.J. & Pitt, M.A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin and Review, 4*, 79-95.

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

Appendix

Let $\mathbf{A} = [a_{ij}]$ be an $n \times m$ design matrix indicating the availability of depth cue j in condition i. Specifically, if cue j is present in condition i, then $a_{ij} = 1$, otherwise $a_{ij} = 0$. The LIM is defined by the matrix equation, $\mathbf{y} = \mathbf{Ax}$, where $\mathbf{y} = [y_i]$ is an $n \times 1$ matrix of observations and $\mathbf{x} = [x_j]$ is an $m \times 1$ matrix of scale values. Using this notation, the FLMP is defined as,

$$y_i = \frac{u_i}{u_i + v_i}$$

where,

$$u_i = \prod_j^m x_{ij}^*$$

$$v_i = \prod_j^m \left(1 - x_{ij}^*\right)$$

and,

$$x_{ij}^* = \begin{cases} x_j, & a_{ij} = 1 \\ \left(1 - x_j\right), & a_{ij} = 0 \end{cases}$$

Note that the estimates of $\mathbf{x}$ will generally be different for the two models.

It may also be noted that the FLMP is a linear model under the logit transform. Let $L(y_i)$ be the logit transform of y_i given by,

$$L\left(y_i\right) = \ln\left[\frac{y_i}{1 - y_i}\right]$$

Then it follows from the definition of the FLMP that,

$$L\left(y_i\right) = \sum_j^m L\left(x_{ij}^*\right)$$

where,

$$L\left(x_{ij}^*\right) = \begin{cases} L\left(x_j\right), & a_{ij} = 1 \\ -L\left(x_j\right), & a_{ij} = 0 \end{cases}$$

Let $\hat{\mathbf{A}} = [\hat{a}_{ij}]$ be the design matrix given by, $\hat{a}_{ij} = 2a_{ij} - 1$, let $\hat{\mathbf{y}}$ be the logit transform of $\mathbf{y}$, and let $\hat{\mathbf{x}}$ be the logit transform of $\mathbf{x}$. Then the FLMP is defined by the matrix equation, $\hat{\mathbf{y}} = \hat{\mathbf{A}}\hat{\mathbf{x}}$.