DISCUSSION

GRUESOME SIMPLICITY*

GRAHAM PRIEST

University of St. Andrews†

1. Introduction. The topic of this paper is one of the main aspects of induction—curve fitting. Given two related quantities x, y, and a finite set of observed corresponding values $S = \{\langle x_i y_i \rangle; i \in I\}$, we have to predict the value y_0 of y for some new value x_0 of x. This is done by choosing the best function (curve) f which "fits" S and hence determining the value y_0 such that $y_0 = f(x_0)$. The problem is of course, to choose the "best" curve from among the non-denumerably many which fit S.

The standard solution is to choose the simplest curve (see e.g., [3], [4]). The characterization of simplicity is a problem, but nearly everyone seems to agree that if we have two families of functions defined by the equations

$$(1) \qquad y = f(a_1, \ldots, a_n, x)$$

$$(2) \qquad y = g(b_1, \ldots, b_m, x)$$

(where the a_i and b_j are parameters) then the members of the first family are simpler than those of the second family if $n < m$, i.e., equation (1) has fewer parameters than equation (2). Thus a straight line defined by the equation $y = a_1 x + a_2$ is simpler than the parabola defined by $y = b_1 x^2 + b_2 x + b_3$.

This account has a number of problems (see e.g., [1]). The point of the present paper is to show that the family of curves defined by an equation of the form of (1) is not invariant under certain very natural transformations—in non-technical terms: which prediction is best depends not on the situation but on how you describe it. (Equivalent descriptions do not give the same answers).

2. A Simple Example. Let us start with a simple case. We observe a moving particle and note its velocity v, and momentum p. It is

*Received November, 1975.

†Current affiliation: University of Western Australia.

Philosophy of Science, 43 (1976) pp. 432–437.

found that when $v = 2$, $p = 6$ and when $v = 3$, $p = 8$. We now ask what the best prediction is for its momentum when $v = 4$. Obviously the curve that best fits the data is the straight line

$$p = 2v + 2$$

and hence we predict that when $v = 4$, $p = 10$.

But now suppose we decide to correlate the velocity with the (classical) kinetic energy of the particle E $(= pv/2)$, computed from the same data. We have that when $v = 2$, $E = 6$ and when $v = 3$, $E = 12$. Again the curve that best fits this data is a straight line:

$$E = 6v - 6$$

Hence we predict that when $v = 4$, $E = 18$. But since $E = pv/2$, $p = 2E/v$, so the corresponding value for p is 9. This is clearly incompatible with our previous "best" prediction.

In other words, our best predictions for the momentum will depend entirely on whether we decide to correlate the velocity with the energy or with the momentum.

3. Generalization. This simplistic argument generalizes as follows. We are given the data $S = \{\langle x_i\, y_i\rangle; i \in I\}$ and asked to predict the value y_0 of y that will correspond to the value x_0 of x. Suppose that f_1 is the simplest curve (according to some criterion of simplicity which embodies the principle of Section 1) that fits the data. Then our predicted value

$$y_0 = f_1(x_0)$$

Now suppose that f_2 is an arbitrary curve which also fits the data (with the proviso that f_2 never takes the value 0). We then have that

(3) $\quad y_i = f_1(x_i)$, and

(4) $\quad y_i = f_2(x_i)$

for all values of i, since both curves fit the data.

But suppose we correlate x, not with y, but with y', where

(5) $\quad y' = y/f_2(x)$

The value y_i' of y' corresponding to x_i which the data gives is

$$y_i' = y_i/f_2(x_i)$$

and using equation (3) and (4), we see that

$y'_i\, f_2(x_i) = f_1(x_i)$, and

$y'_i\, f_2(x_i) = f_2(x_i)$

In other words the curves

(6) $\quad y' = f_1(x)/f_2(x)$, and

(7) $\quad y' = 1$

both fit the data for x and y'.

Clearly (7) will be much simpler than (6) in general. So we accept (7) as the correlation between x and y'. Thus when $x = x_0$, $y' = 1$ and so by (5)

$1 = y_0/f_2(x_0)$, i.e.

$y_0 = f_2(x_0)$

But f_2 was arbitrary.

4. Another Example. A similar situation can be obtained in a different way. Let S, f_1, f_2 be as in Section **3**, except that now f_2 is an arbitrary single valued function that fits the data. Hence it has an inverse function f_2^{-1} such that

$$f_2 f_2^{-1}(x) = f_2^{-1} f_2(x) = x$$

Now instead of correlating x and y we will correlate x' with y where

(8) $\quad x' = f_2(x)$

The value x'_i of x' corresponding to y_i given by the data is

$x'_i = f_2(x_i)$

Thus since $f_2^{-1}(x'_i) = x_i$ we have from (3) and (4)

$y_i = f_1 f_2^{-1}(x'_i)$, and

$y_i = f_1 f_1^{-1}(x'_i)$

Thus the curves

(9) $\quad y = f_1 f_2^{-1}(x')$, and

(10) $\quad y = x'$

fit the data for x' and y.

Since (10) will clearly be simpler than (9) in general, we accept it as the correlation between x' and y. But when $x = x_0$

$x' = f_2(x_0)$ by (8), and so

$y_0 = f_2(x_0)$ by (10).

As an example of this sort of situation, if y is the length of a moving body and x its velocity, then x' could by $mx\,(1 - x^2/c^2)^{-1/2}$, which is its relativistic momentum.

We can summarize the general situation thus: if f is the curve from the simplest family which fits data S and if θ is virtually any transformation of the cartesian plane into itself, then the image of f under θ will not in general be the curve from the simplest family which fits the image of S under θ. Thus an appeal to simplicity will not help us with this problem of induction.

5. Grue Rears its Ugly Head Again. In this section of the paper, I want to show that the above observations are relevant to another standard puzzle of induction: the grue paradox (see e.g. [2]).

The situation is as follows: we have some emeralds which we have observed every morning for the last n days. (Let us suppose that $t = 0$ now and observations were made at $t = 0, -1, \ldots, -(n-1)$).

Each morning we have measured the frequency ν of the emitted light rays. This has always been g, a frequency in the green part of the spectrum, to within experimental error. We now ask what ν will be at $t = 1$ (tomorrow).

Clearly, if we plot ν against t, then the "best fit" is the straight line $\nu = g$. Hence we predict that the emeralds will be green.

But now let b be a frequency in the blue part of the spectrum and let $f(t)$ be (the principle value of)

$$(b + g)/2 + ((b - g) \arctan \omega\,(t - 1/2))/\pi$$

To save the reader working out what this looks like, fig. 1 is a sketch of the curve. The function has asymptotes $\nu = b$ and $\nu = g$, and flips over between $t = 0$ and $t = 1$. We can choose ω such that for $t < 0$, $g - f(t)$ is as small as we please and for $t > 1$, $f(t) - b$ is as small as we please (and the greater ω, the faster the flip over). Suppose ω is large enough to make these quantities less than experimental error.

Now if instead of plotting ν against t we plot $\nu' = \nu/f(t)$ against t, then for $t < 0$, the values of ν' given by the data are all equal to 1, to within experimental error. Hence, as in Section **3**, we predict that when $t = 1$, $\nu' = 1$, i.e., $1 = \nu/f(t)$. So when $t = 1$, $\nu = f(1) = b$, to within experimental error, i.e., the emeralds will be blue.

Thus we see that the grue paradox is just a special case of our more general problem. It follows that all solutions of the grue paradox along phenomenological lines or in terms of the learnability of 'grue' (amongst others) can not be correct.

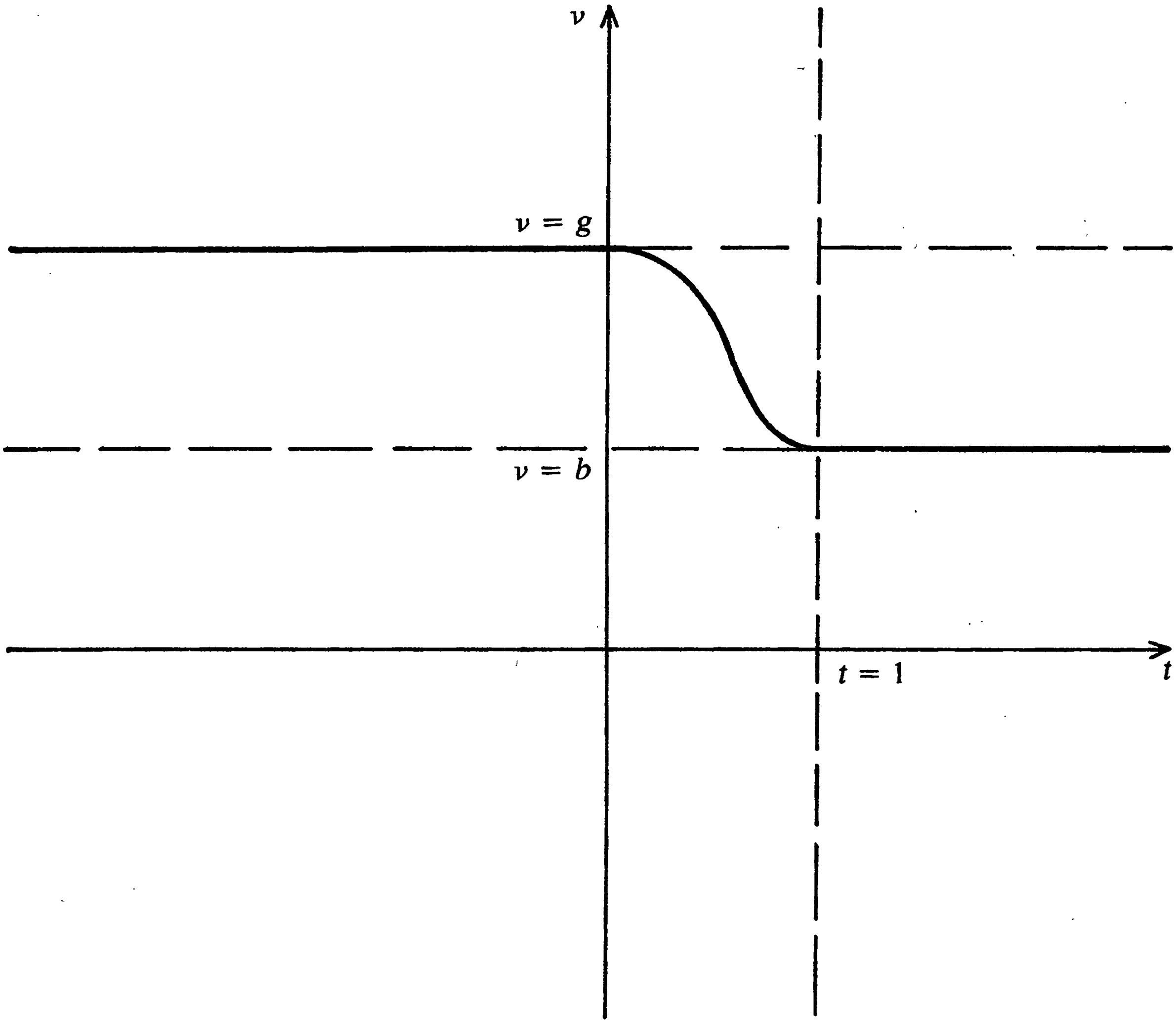

Fig. 1.

6. Conclusion. I do not feel in position at present to suggest a resolution of these problems. They seem to be fundamental to our intuitions about induction and not susceptible to a quick resolution.

The problem is of course why we should correlate the quantities x and y rather than the related quantities x' and y'. Rarely, both procedures may be natural, in the sense that all the quantities are already recognized as having theoretical significance (as in the example in Section 4). In this situation the reasonable thing to do would be to run a test to eliminate one of the hypotheses.

However, in the (uncountable) majority of cases, x', y' will be artificial in the sense that they have (as yet) no recognized theoretical significance. This I think is the heart of the matter. Especially in view of the fact that the quantities that have theoretical significance change in time (e.g., $mv(1 - v^2/c^2)^{-1/2}$ had no particular significance in classical dynamics, but it has in special relativity), why if at all should we discriminate in favor of those quantities which have theoretical significance for us?

There is no objective form of simplicity one can appeal to. (For example, even through in Section **5** v' looks more complicated than v, we can of course define v in terms of v' making v look complicated). Thus the only sense in which x, y will be simpler than x', y' is the entirely subjective one of familiarity.

Thus the question is, what price theoretical significance and familiarity?

REFERENCES

[1] Ackermann, R. "Inductive Simplicity." *Philosophy of Science* 28 (1961): 152-161.
[2] Goodman, N. *Fact Fiction and Forecast* Ch. VII Cambridge: Harvard University Press, 1955.
[3] Kemeny, J. "The Use of, Simplicity in Induction" *Philosophical Review* 62 (1953): 391-408.
[4] Popper, K. *Logic of Scientific Discovery* New York: Basic Books, 1959. Esp. Ch. VII.