Mathematical Modelling and Contrastive Explanation

ADAM MORTON
University of Bristol
Bristol BS8 lTB
England

This is an enquiry into flawed explanations. Most of the effort in studies of the concept of explanation, scientific or otherwise, has gone into the contrast between clear cases of explanation and clear non-explanations. (Controversial cases are to be put into one box or another.) My interest is rather different. I want to discuss explanations which are clearly imperfect, but also clearly not completely worthless as explanations. Sometimes they are the best explanations we can get of some phenomena. My object is to find the right vocabulary for discussing their degree or character of imperfection. My interest in these questions comes from an interest in commonsense psychological explanation, but that will not feature here. Instead I shall discuss mathematical modelling.

There is an enormous range of things that can be called mathematical models.[1] Sometimes a mathematically expressed theory is called a mathematical model to indicate agnosticism about its physical significance. Sometimes what is called a mathematical model is just a rather complex database, imposing a structure on a body of observations. I shall focus on one particular class of scientific activities. The activities that interest me involve the use of a mathe

matical formalism, the model, tied to a theory in a particular way. There are two main features of this. First, the whole function of the model is to derive explanations and predictions that the theory alone cannot give. But alone the model has no explanatory force and does not receive any confirmation from its explanatory successes. I call this *subsidiarity*. Second, crucial features of the model are arbitrary in a way that makes it hard to give them physical significance. I call this *inhomogeneity*. (More about both of these below, and examples. Some of the examples only fit within my description if you believe some of the things I am arguing for.) My first aim is to show that there is something very interesting going on with this kind of modelling, which complicates our picture of science in an interesting way. My second aim is to find a way of expressing what is successful and what is deficient about the explanations these mathematical models provide. In doing this I introduce a distinction between the width and depth (or scope and force) of explanations, based on the Dretske-Garfinkel idea of contrastive explanation. (That is my way of explaining the distinction. But I suspect that it could be sustained independently of that foundation.)

I Inhomogeneity

The kinds of mathematical modelling I am interested in are marked by the two features, 'inhomogeneity,' and 'subsidiarity.' The first of these is easiest to explain. Here are some examples of it.

One typical mathematical treatment of turbulent fluid flow consists of a set of partial differential equations involving some dozen parameters.[2] Given the right values for the parameters the model can predict the behaviour of a fluid in some quite complicated circumstances, for example when it flows in a pipe one side of which is considerably rougher than the other. It is the parameters that matter

here. They are arbitrary 'system parameters/ and do not include 'control' parameters defining the system being modelled, for example the viscosity of the fluid and its initial rate of flow. (Some of the parameters may be redundant: no one can 'close' the equations so as to characterise the system parameters in terms of the values of the functions the equations define.) For each range of values of the control parameters there are values of the system parameters for which the equations give good predictions. When the behavioral parameters pass crucial thresholds the predictions are no longer good and new parameters have to be fixed. There is no formula for getting suitable values of the parameters. In fact there is no assurance that the values providing accurate predictions are unique or that there are not other values giving more accurate predictions.

This is what I call inhomogeneity. For every range of the behavioral variables there is a suitable set of system parameters, but even slightly different values of the behavioral variables may call for very different values of the system parameters. They jump around wildly. This makes it difficult to take the system parameters to represent attributes of the physical system at hand, unless there is some good reason to believe that fundamental features of the system undergo drastic changes at these thresholds. (A different model might employ parameters which had to be changed at quite unrelated thresholds.) Moreover, the system parameters may be nonunique or redundant. One knows neither whether there are better but radically different values for them, giving equally good predictions, nor whether a given combination of values is consistent with the equations.

(A distinction is relevant here which I will say more about later. The model may be taken as a direct description of the phenomena and their causes. Or it may be taken as an approximation or manageable version of some other theory which, while giving a physically real and more complete description, does not lend itself to explanation and prediction. For example a model of fluid flow can often be taken as a manageable substitute for the Navier-Stokes equations. In this case the more ultimate theory may give reason for believing that some of the inhomogeneities of the model do in fact correspond to sudden fluctuations of underlying quantities. Very often, though, it will not.)

The second example comes from economics. One often models the choices of economic agents by postulating a typical utility function, allowing cardinal comparisons between agents' preferences among simple options and gambles within a given area such as choices of given commodities or the balance between work and leisure. The modelling often gives a good fit to present and future data. But there is an inhomogeneity here too. To handle one bit of choice behavior one postulates one utility function, and to handle another one postulates another. There is no assurance, and in fact usually no attempt, to form a consistent picture of the overall utility functions of economic agents. (That would be getting too near to psychology.) And my impression is that the few attempts there are to explain rather different economic choices within a single attribution of cardinal utility functions-e.g. the propensity to buy insurance and the (opposed) propensity to speculative investment[3]-are generally thought by economists to be misguided.

A third example is provided by catastrophe theory. Perhaps more precisely by Zeemanism (known in France as 'Thomisme'!) which I take to be the ambition to explain just about everything in sight by catastrophe theory. The procedure is this: one has a phenomenon which involves discontinuous and hard-to-predict transitions of a physical system from one state to another. One then models this by representing the state of the system by the value of a function which when mapped against the values of some 'control' parameters produces a folded surface. The system can then be thought of as dropping over the edge of the folds of this surface from one equilibrium to another, at crucial moments of transition. Catastrophe theorists provide models along these lines for no end of phenomena: the development of embryos, the capsizing of ships, the bending of beams, changes of mood, prison riots, anorexia nervosa

The essential mathematical move is that it must be possible to interpret the folding surface in question as the set of extremal

points of a potential energy function, so that the resulting catastrophes-the patterns of discontinuous transition-can be classified in a very deep and powerful way due to Rene Thom.[4]

The inhomogeneity here is rather like that in the economic example above. There are typically many functions from control parameters to behavioral states which have the right folds to generate the observed catastrophes. The system is successfully modelled as long as one of them is found. But if a slightly larger range of control variables is considered, or the behavior of a larger or slightly variant system is considered, a quite different function may be needed. Successful modelling does not require that the function used be stable under extensions or variations.

A fourth example is a bit different, in that instead of a standard bit of mathematical modelling it uses a philosophically controversial account of physics. One of the main arguments used by Nancy Cartwright, in *How the Laws of Physics Lie,* to argue that the formalism of quantum mechanics should not be taken as a body of claims about the physical structure of things, is in effect a claim of inhomogeneity. Her argument centres on the choice of functions representing crucial physical quantities of a physical system, notably its total energy. She claims that while very often we can choose functions which allow us to get the right answers, quantum mechanics does not tell us how to choose them from the many reasonable candidates. To quote her (and her quoting Merzbacher):

> In quantum mechanics the correspondence principle tells us to work by analogy with classical mechanics, but the helpfulness of this suggestion soon runs out. We carry on by using our physical intuitions, analogies we see with other cases, specializations of more general considerations, and so forth. Sometimes, we even choose the models we do because the functions we write down are ones we can solve. As Merzbacher remarks about the Schroedinger equation: Quantum dynamics contains no general prescription for the construction of the operator H whose existence it asserts. The Hamiltonian operator must be

This observation is certainly right about quantum mechanics at some stages of its development. It probably overestimates the uncertainty there is in the choice of a Hamiltonian nowadays, given both the accumulation of experience about what assumptions prove to be mathematically sustainable and the development of a tradition which specifies what is to count as a suitable quantum mechanical description. The result is that if this tradition (what I below call a cookbook) is counted as part of quantum mechanics, then there is no great degree of inhomogeneity.

(It is not at all clear-to me, at any rate-what contrasts this makes with classical mechanics. There too to get an account of a system we have to supply, for example, forces, initial and boundary conditions, and a formula for the potential energy of the system. And the formalism does not give any of these to you on a plate. So there is room for the same kinds of inhomogeneity there too. But in practice there seems to be much less of it. The reason seems to lie in two things. First it seems easier to link smaller to larger systems, so that going from a component of a complex system to the whole system is a smoother business. This may be a consequence of the second difference, that physical intuition and the tradition of physics specifies more exactly what form the Hamiltonian must take. For both these reasons one can more easily take one's characterisation of a system as representing real properties of it. On the other hand in all real cases there are boundary constraints, and these too are often formulated on the basis of 'physical intuitions, analogies we see with other cases, specializations of more general considerations, and ... because the functions we write down are ones we can solve.' So, to this extent, the classical formulation varies unsystematically from situation to situation. And though this effect is quite slight in

the cases to which classical mechanics happily applies, one reason it won't go away is that the truth about nature, including the truth about what happens at the edges of systems, is not classical.)

II Subsidiarity: Theories, strategies, and cookbooks

Inhomogeneity occurs very naturally in some scientific contexts. In fact, it is sometimes quite advantageous. Let me describe its natural habitats.

The simplest context for it is an existentially quantified theory. , Mechanics says that every particle has a mass and a position, and every system has a Hamiltonian function, but does not say what they are. Microeconomics says that every agent has an utility function giving cardinal comparisons between, e.g., different amounts of money, but does not specify it. Such a theory will not by itself have many observable consequences. To get explanations or predictions out of the theory one will have to specify values for numerical and function variables. Such a specification is the simplest case of a mathematical model. And when the specified values cannot themselves be directly measured and vary from case to case in a way that the theory cannot explain, the model is a separate entity from the theory. It will typically vary while the theory remains constant.

(Note that this is not true of, e.g., masses of particles. Conservation laws guarantee that. And it is controversial whether it should be allowed for, e.g., utility functions. My point is only that when the values do vary from application to application we have a mathematical model that is significantly different from the theory it supplements.)

Many theories require more than a simple filling in of values in order to connect them with observable data, or with a particular body of data. Very often quantities must be postulated which are not mentioned in the theory, and new relationships between quantities must be postulated. The turbulent flow example is an instance of this, if we take the background theory to be the mechanics of incompressible fluids and if we take the model as simply specifying more quantities which allow predictions to be extracted from it. (But see the 'other case' at the end of this section.) Then the model which augments the theory has more of the appearance of a theory in its

own right. But there are two reasons for seeing it as something other than a regular theory. First there is its different epistemological position, being tied for support and intelligibility just to one larger theory rather than to a whole area of science. (This is a matter of degree. If the mother theory is large and diffuse this factor clearly does not produce an important contrast.) Then there is inhomogeneity, of course. The values of functions and parameters in the model will vary from application to application. (So in fact the model would not be redescribed as a single theory at alt but as a cluster of theories, or as an existentially quantified theory plus a cluster of value-specifying mini-models.)

Models of either of these two kinds are often used to test theories. Very often a theory will lack the connections with experimental data (or with a particular appealing source of data) which would provide tests for it. Then one often constructs a model specifying more and postulating more, in the hope of matching the data. There is no claim that the values postulated in the model are the true ones. (Sometimes there is no claim that the functional relationships have any causal significance.) If such a model can be found, the theory receives some, fairly weak, confirmation. And if no such model can be found – all plausible values for variables and additional functional relationships lead to the wrong numbers-that is clearly quite bad news for the theory.

(There is an interesting asymmetry here. If true predictions are forthcoming, the theory takes much of the credit. It fits reality at least well enough to allow the construction of a model. But if false predictions are produced, the first object of blame is the model. The only case in which the theory cannot escape blame is when all attempts to construct a prediction-producing model fail.)

The values specified in a model are rarely just plucked out of the air. The theoretical background is usually part of a scientific tradition or research programme. (There is typically a nested structure of research programmes, ranging from the immediate theoretical project to platitudes of scientific respectability.) And this often gives a fairly specific strategy for constructing models to account for the behaviour of particular kinds of system, leaving a larger or smaller amount up to the ingenuity or intuition of the theorist. I call this strategy the cookbook.

The cookbook very often adopts a realist attitude, specifying the way the model may be constructed in terms of the objective structure and causal construction of the system to be modelled. Textbooks of mathematical modelling discuss different strategies for getting mathematical treatments of systems of different physical types and indicating both the form the model should take and the general patterns of mathematical results and techniques (the 'mathematical phenomena' as M. V. Berry calls them) which often work to get useful data out of the model. [6] The cookbook for quantum mechanics says (or rather, begins) 'look the corresponding function for a classical system with the same physical structure as the system you are studying.' The cookbook for catastrophe theory begins 'try to describe the system in terms of variables which can be divided into two sets, control variables and behavior variables, such that the relation between them can be interpreted as a set of equilibria of an underlying dynamical system in such a way that the qualitative behavior of the system can be characterised as one of the standard catastrophe-shapes.' In neither of these cases does the cookbook tell exactly how to go about setting up the model. And in both of them it gives no general assurance that the values we invent to get a best fit with aspects of the same or related systems will fit together in any homogeneous way. That is the way it is generally.

A cookbook can exist without a theory. The most interesting cases of this are those in which the strategy for constructing models requires that a model be backed up by a theory but is fairly neutral about the content of the theory. This is the case with the cookbook for catastrophe theory. It requires that the behavior of the system be the product of an underlying dynamical system. But that does not mean that it has to consist of particles moving according to classical mechanics. Rather, it means that the mechanisms underlying the behavior must be produced by some causal processes similar to mechanics in only a very abstract way. Then the strategy for filling a theory out with a model has become autonomous, requiring

the presence of a theory but consistent with an indefinitely wide variety of theories.

The central case here is that in which there is a definite theory which leaves some crucial things unspecified, and so needs to be supplemented with a model. (Then a cookbook may come in, to say what kinds of supplementation are allowed.) The model is then clearly subsidiary to the theory. And clearly indispensible.

There is also a very important other case, equally central as an example of mathematical modelling, which should be placed alongside the first. In this other case, there is a single completely specified theory. But it does not lend itself to making predictions. The usual reason is just that we do not have a general solution to the equations. Then too we can profitably construct models, which bridge the gap between the theory and the phenomena. The model is generally a simpler set of mathematical conditions than those implied by the theory, whose consequences are easier to calculate. Either the equations can be solved, or approximate solutions are easier to get than for the full theory. (Or, a more modern form, the model just is a programme for computing the consequences of assumptions in a way that is not too much at variance with the main theory.)

A model of this sort is related to its theory in rather a different way than a model of the first sort is. Since it is a simpler or more manageable version of the theory, it may not even be consistent with it. What is required of it is that it have roughly the observational consequences that the theory would, under the given conditions. Or, more cautiously, that it allow us to make a stab at formulating the observational consequences the theory would have under particular conditions. (This is something that needs a lot more study: the extraction of predictions from a theory by means of a simplification which is actually inconsistent with it.)

But models used to tame an unsolvable theory are epistemically much like models used to complete an underspecified theory. They also allow tests of a theory which cannot be tested by itself. And the model's justification is entirely in terms of its power to set up such tests. (Unlike the theory which is justified in part in terms of its connections with the rest of science.) Models of both kinds have the peculiar epistemic status of mediating the flow of evidence with-

| | | |

out accumulating it for themselves.[7] To the extent that the model is not simply a special case of the theory, but rather makes further assumptions justified only by their prediction-extracting power, and specific only to the explanation at hand, they are not taken to describe the ultimate causes of the phenomena in question. Or even to be claims to truth.

III An invented example

Here is an invented example that brings out some of the points I have been making. Consider some data; think of them as outputs of a physical system. The system has one input, i, and given this input it produces in succession three outputs, 01, 02, 03. Considering these outputs as the values of a function O(t), t ranging from 1 to 3, the data are:

	output O(t)		
input i	t=1	t=2	t=3
1	2	3	2
2	4	3	2
3	6	4	2
4	8	4	2

These data can be captured by a simple formula with two parameters:

$$O(t) = (2-t)(3-t)i + (2 - (2 - (2-t)(3-t)A + (t-1)(3-t)B$$

No values of A and B will fit all the data. For t = 1 and 2 the data are caught if A = B = I, for t = 3 and 4 A = 1 and B = 2.

Other formulas can, of course, fit the same data. Here are two such, the first with two parameters and the second with one parameter:

P(t) = (2 - t)(3 - t)i + (t -1)(3 -l)C + (t -l)(t - 2)0   Q(t) = (2-t)(3-t)i + (t-l)(3-t)E + (t-l) (t-2)

Given that 0, P, and Q all fit the data, is there anything to choose between them? Yes: two things in particular. First, there is potential explanatory force. 0 could explain, but P could not, why the pattern of data is always '2t, then up then down.' It follows from 0 that the data will have this qualitative pattern, whatever values A and B take. But this does not follow from P.

(1 say 'potential explanatory force' and 'could explain' because the derivation from the equation only explains the data if one has some reason to believe that the equation bears some relation to the reasons why the data take the form they do. 1 return to this point in the next section.)

Second, there is extendability to a larger range of data. Suppose the data continue, for i = 5,6,7,8 as follows:

| | output O(t) | | |
input i	t=l	t=2	t=3
5	10	6	4
6	12	6	4
7	14	7	4
8	16	7	4

These further data conform to 0 given that for i = 5 and 6 A = B = 2, and for i= 7 and 8 A = 2 and B = 3. But they cannot be brought under Q, for any values of E. So if we want to catch

all the data with a formula that entails that the pattern '2t, up, down' is intrinsic to them, 0 is preferable to both P and Q.

At this point there are three possibilities. 0 may be just a convenient summary of the input/output relationship. (In which case the claim to have explained the '2t, up, down' pattern is pretty hollow, so P is as good as 0.) Let us suppose that this is not the case. Then there are two possibilities. 0 may be the result of a cookbook for modelling phenomena of some particular kind, which gives general guidelines which given the details of the particular case entail O. Or 0 may be entailed by a theory about the structure and behavior of systems of the type in question. The basic theory will be or entail an assertion of the form

T   (t)∃a∃b (O(t) = (2-t)(3-t)i + (2 - (2 - (2-t)(3-t)a +   (t-l)(3-t)b)

T does not entail any data. Its existentially quantified form is compatible with too many possibilities. Still, it may be a true account of the system in question. But if we want to test it or explain data with it, we have to replace the existential quantifiers with something more specific. The way to do this is to add specific hypotheses about the values a and b take for particular values of t. Then we have P, taken as including the relations between t, A, and B as specified above. Thus expanded, T is tested by and explains the data. (Note that tracing Q back to a theory in the same way would yield an existentially quantified assertion which does not include the data among its models.)

Why not include the values for A and B in the theory? Two related reasons. T will typically apply to other systems, and the values for A and B may work just for the system at hand. And T may be derived from, or otherwise linked to, larger bits of theory, which explain or motivate T in its existentially quantified form but give no reason to suppose that the values of the variables are as the model supposes. So we are best off keeping theory and model, T and 0, separate.

There is not a clear line between the second of the two possible interpretations of O-that it is the result of a cookbook about how to model systems of various kinds-and the third-that it is the result of a theory about the structure of systems of that kind. For a

modelling cookbook can be thought of as a higher-order theory: systems of this kind are such that models of these types match their behavior. (They are not theories about the structure or causal propeties of systems, though.) So if you take a model produced by a cookbook as a special case of a model produced by a theory the conclusions I just drew still hold. Keeping the model separate from the 'theory' allows tests without committing the theory to unintended specificities, and broadens the scope of the theory's application while keeping its main principles unchanged.

What is a higher-order theory? Differential equations are higher-order theories, in that they do not say, e.g., 'the particles follow path A,' but 'the path of the particles is given by a function p satisfying equation E.' And typically there are many solutions to E, so that to pick out the right one, boundary and initial conditions are needed. So a potentiality for nonhomogeneity is built into one of the most basic prescriptions of our scientific culture- 'say it with differential equations.' So we may be deeply biased to a division of labour between theories and mathematical models. Note that the relevant existential quantifiers here are typically second-order: they assert the existence of functions rather than of numbers.)

IV Constrastive **explanation**

Mathematical models can often be used to explain things. In my unreal example 0 can explain the eight triples of numbers representing the system's output. 0 can also explain why the pattern in all eight cases is '2t, up, down.' So there is both a qualitative and a quantitative element to the explanation. The real cases I cited earlier are similar. And the example of quantum mechanics shows that a very great explanatory power can be combined with a considerable inhomogeneity. So, many explanatory claims arise. The theory is tested by the data (by use of the model) but does not predict it. The theory explains some qualitative features of the data but not its exact quantitative values. The theory-plus-model explains and is tested by the exact values.

But these explanatory claims are tricky. Taken at face value they would seem to legitimate all the wilder explanatory claims of catas-

trophe theory. (And they seem to deal too quickly with natural doubts about the explanatory force of highly inhomogeneous models such as the turbulence model.) My strategy for sorting out these questions is to try to deflect attention away from questions of the form 'what is the strength-or value-of the explanation of 0 by M' to questions of the form 'what aspect of 0 can M explain?'

In effect this is to take over Dretske and Garfinkel's idea of contrastive explanation.s The idea is to see explanations as saying not why something happened but why it happened in one way rather than another. So every explanation is made in the context of a contrast space: the explained event is contrasted with a set of others which might have happened but didn't, for reasons which the explanation makes clear. One example of this is implicit in the opposition between quantitative and qualitative explanations: to explain a quantitative phenomenon is to explain why some observable took on the values that it did, against a contrast space of other possible values, while to explain a qualitative phenomenon is to explain why a 'pattern of observables was found, against a contrast space of other possible patterns. But since my interest is in flawed explanations I shall also consider cases where there is an explanation but it is not fully contrastive.

Consider some contexts for some explanations. I shall use two examples, catastrophe theory and models like the artificial example of the last section, which I shall call 'parameter models.' Catastrophe theory because the explanations it gives are notoriously flawed, and parameter models as a substitute for the fluid dynamics case, which would get very technical. In each case the important thing to note is for what attributes A is it explained why the state of the system does *not* have A.

The aim of catastrophe-theoretical explanations is to give an explanatory hold on the qualitative aspect of a phenomenon-why discontinuous transitions take the pattern that they do. Often there

is a further ambition, to explain the quantitative aspect- why transitions occur at the times and places that they do. There are several grades of this. The lowest (grade zero) is when there is no reason to believe that the folding surface represents equilibria of an underlying mechanical system. Then there is no explanation of anything. We just have a rather suggestive database. We cannot explain either why the system evolves as it does or why it does not evolve in some other way.

The analog of this with parameter models occurs when although there seem to be values of the parameters for each selection of data for which the formula entails the right number, there is no reason to believe that there is any physically significant function which has the 'right' values of the parameters as values. Then too we have little more than a database.

The next grade (grade one) of explanatory force comes in catastrophe theory when we can assume that there is an underlying dynamical system whose equilibria are topologically like the surface postulated for the data. Then we have an explanation of the qualitative aspect, of why the catastrophe has the form that it does. And we can explain why the catastrophe does not take another form. But we cannot explain why it occurs when and where it does.

The analog of this for parameter models occurs when we can assume that there is a physically significant function giving the right values of the parameters, though we cannot specify it. Then we can explain quantitative aspects of the data (e.g. the '2t, up, down' feature), and explain why they do not take other forms. But we cannot explain why they have the numerical values they do, and not others.

The top grade (grade 2) comes for catastrophe theoretical explanations when we are given an explicit link to an explicit specification of the system's dynamics, determining the surface as that of extremal points of an energy function. Then we have an explanation both of the form of the catastrophe and of why it occurred at just the points in time and space that it did, and not others.

And the analog of this for parameter models occurs when there is a physically significant function determining the values of the parameters, and we can specify it. Then we can explain both why the data have the values that they do, rather than other values, and why they fall into the patterns that they do, rather than other patterns.

These are not the only possibilities. Two in-between possibilities are important. There is a grade between grade zero and grade one, in which, although we cannot explain why the data have the qualitative aspect that they do, rather than some other aspect, we can give some sort of a (non-contrastive) explanation of the pattern that they take. This would be the case with a catastrophe theoretical explanation in which, although we have no assurance that the surface represents the equilibria of the system, we believe that we are dealing with a mechanical system and that it follows in some way from its operating principles that the catastrophes will take the form that they do. Then the explanation matches those of the second grade in width, though , not in depth. That is, we can give an explanation which gives a reason why the catastrophes take the form that they do, but the explanation does not give a reason why they do not take a given different form. They lack depth, contrastive force. To have contrastive force the explanation would have to more explicit about the connections between the surface and a fuller mechanical description of the system.

The parameter model analog occurs when although we have no reason to believe that the parametrised formula is an accurate representation of anything causally relevant to the data, we do have reason to believe that some formula of the same general form can be found, which does have causal relevance. Then we may be able to explain why the data exhibit some patterns without explaining why they do not exhibit others.

The other intermediate possibility lies between grades one and two. It occurs when we can get a contrastive explanation for the qualitative aspects of the data but only a non-contrastive explanation for the quantitative aspects. This occurs with a parameter model (to give that case first this time) when we have reason to believe that there is a physically significant function which gives the values of the parameters, but we do not know what it is. Then we have a grasp of the physical process behind the data, which we can state in a form that entails that they take this form and not another, and which moreover give the causal reasons why they take

This form and not another, and which moreover give the causal reasons why they take the numerical values that they do. But it does not allow us to explain why they do not take on different values.

The catastrophe theoretical analog of this occurs when we can reasonably postulate that the surface is that of the extrema of a mechanical system (whose evolution accounts for the behavior observed) but cannot explicitly characterise it. Then too width and depth, scope and force, come apart. We can explain why the catastrophes occur when and where they do, as well as the form they take-just like the second grade-but we cannot explain why they do not occur at other points. Lack of contrast.

The width/depth distinction which these intermediate cases press on us seems to be part, at any rate, of the diagnosis of many cases in which mathematical modelling yields explanations which seem at once apt and flawed. Go back to the very first example, that of modelling turbulent fluid motion. The most puzzling case is that in which we believe that although the (arbitrarily varying) parameters do not themselves represent anything real in the physics of the system the pattern of fluid flow they entail is a consequence of whatever the true underlying principles are-so that, for example, there are most likely choices of values for the parameters which will extend the applicability of the equations to cases beyond those it has presently been applied to. Then we have an explanation of grade *1112* above. That is, we believe in this case, that for every range of the control variables and boundary conditions (velocity, viscosity, shape of pipe) there is a set of values for the system parameters for which the equations accurately describe the flow. Then one gets a weak explanation of the values of quantities describing the flow. It is weak because although it explains why these quantities have the values they do, one cannot give reasons why the parameters have the values that are necessary to get the explanation to work. Therefore one cannot explain why other values are not found. Width is bought at the price of depth.

On the other hand, in cases like this qualitative features of the phenomena may be explained in a way that is independent of the choice of parameters. (Remember how the '2t, up, down' pattern in the invented example was independent of the choice of A and B.) So these can-sometimes, for qualitative aspects that are independent of the numerical values - be given fully contrastive explanations.

(In fact, the possibility of getting satisfactory qualitative explanations when quantitative explanations are unobtainable or problematic is one of the main appeals of catastrophe theory and of a large and

developing part of mechanics of which it is a part. I believe that in the study of chaotic systems, another part of 'qualitative mechanics,' one can have explanations which are fully contrastive for quantitative aspects of a phenomenon and deficient in contrast for qualitative aspects. But that would take more argument.)

V To end: A glimpse **in another direction**

Theory really is very different from observation. There is usually a considerable gap between one's beliefs about how things are structured and what makes them behave, on the one hand, and on the other hand their observable behaviour. Gap-fillers are needed, and mathematical models are one way of filling one kind of gap.

Different gaps need different fillers. If we go from physical science to commonsense psychological explanation things are inevitably rather different. But it is remarkable how much is similar. There is a very general set of background assumptions. (Taken by some to be a very general theory and by others as something rather less explicit and discursive.[9]) And the gap between this and our observations of one another's actions is filled with a shifting and improvised pattern of ascriptions of beliefs, desires, moods, and all the rest, to particular people at particular times. These ascriptions are highly inhomogeneous - we change them as we need to in order to make actions intelligible - and are based on a traditional cookbook, in part culturally transmitted, telling how to imagine (mentally model) another's state of mind. Put like that, brushing out most of the detail, folk psychology and mathematical modelling seem remarkably similar.[10]

They certainly present similar problems of explanation. In folk psychology, too, the inhomogeneity is connected with our often being able to explain 'qualitative' aspects of behaviour (why two lovers quarrel) in ways that do not depend on or generate explanations of 'quantitative' aspects (why they quarrel at that particular moment, why they quarreled three times yesterday). And here, too, this asymmetry is connected with a distinction between the depth and the width of explanations. For the explanation of why two people quarrel in a sense is an explanation of why they quarrelled at some particular time. (They quarreled at 2pm on Thursday the first of March because one resents the other's bossiness and the other can't help teasing the first.) But this doesn't explain why they quarrelled at that moment rather than some other. It is an explanation of grade 1 ½: width has been bought at the price of depth.[11]

1 To appreciate the variety of things that can be called models in physics. See Michael Readhead, 'Models in Physics: *British Journal for the PhIlosophy of Science* 31 (1980) 154-63.

2 See Gordon Reece, A Generalized Reynolds Stress Model of Turbulence, PhD thesis, Imperial College, University of London (1977); D. C. Leslie, *Developments in the Theory of Turbulence* (Oxford: Oxford University Press 1973), especially ch 13; and for the background L. Landau and E. Lifshitz, *Fluid Mechanics* (London: Pergamon 1969).

3 See Milton Friedman and Leonard Savage, 'The Utility Analysis of Choices Involving Risk: *Journal of Political Economy* 56 (1948) 279-304, and Angus Deaton and John Muellbauer, *Economics and Consumer Behavior* (Cambridge: Cambridge University Press 1988).
6 For example, J.G. Andrews and R.R. McLone, eds., *Mathematical Modelling* (London: Butterworths 1976).

4 See Tim Poston and Ian Stewart, *Catastrophe Theory and its Applications* (London: Pittman 1978); Vladimir Arnold, *Catastrophe Theory* (Berlin: Springer 1984), is very eloquent about the wildness of wild applications of the theory; and Christopher Zeeman, 'Catastrophe Theory: *Scientific American* is very stimulating about the line between explanatory and non-explanatory uses of it.

5 Nancy Cartwright, *How the Laws of Physics Lie* (Oxford: Oxford University Press 1983)

7 Colin Howson and Peter Urbach showed, to my surprise, that this is explainable from a Bayesian point of view. Suppose the prior probability of the theory is fairly high, that of the model zero, the probability of the data conditional on the conjunction of theory and model highish, and that of the data conditional on the theory alone zero. Then on conditional is at ion the conjunction of model and theory will rise in probability although that of the model itself will stay at zero.
8 See Fred Dretske, 'Contrastive Statements: *Philosophical Review* 82 (1973), and Alan Garfinkel, *Forms of Explanation* (New Haven: Yale University Press 1981). See also 92-5 of Dretske's *Explaining Behavior* (Cambridge, MA: M.l.T. Press 1988).

9 See Adam Morton, *Frames of Mind* (Oxford: Oxford University Press 1980); Robert Gordon, *The Structure of the Emotions* (Cambridge: Cambridge University Press 1988), for views I approve of, and for a balanced picture see the papers in Radu Bogdan, ed., *Mind and Common Sense* (Cambridge: Cambridge University Press forthcoming).

10 I develop some different similarities between folk psychology and qualitative mechanics in 'the inevitability of folk psychology,' in Radu Bogdan's anthology cited in n. 9.

11 I have been greatly helped by conversations with David Hirschmann, David Papineau, and Gordon Reece. Audiences at the Oxford Philosophy of Science club, and Cambridge HPS seminar, and at LSE provided doses of friendly skepticism. The CJP referees' comments were some of the most helpful I have ever had