Philosophy and the Departments of Knowledge

Frank Jackson
Chinese University of Hong Kong, 13 March 2006

Introduction

What is our world like? What's in it, and what properties are instantiated? Ask at the physics department and you will receive an answer in terms of electrons, photons, force fields, quarks, the curvature of space-time, and so on. Ask at the chemistry department and you will receive an answer in terms of valency, acids, crystal structure, and so on. Ask at the psychology department and will you will receive an answer in terms of motives, language acquisition, perceptual deficits, pain tolerance, emotional intelligence and so on. Ask at the biology department and you will receive an answer in terms of cell membranes, evolutionary forces, carbon-fixing, CO_2 uptake, and so on. Ask at the international relations department and you will receive an answer in terms of the role of the United Nations, the distribution of trouble spots around the globe, national rivalries, and so on. And so it goes. The famous explosion of knowledge has been accompanied by a dramatic increase in specialisation or departmentalisation. Knowledge comes to us in department-framed parcels, wrapped up in the conceptual apparatus distinctive of the various departments.

In this lecture I want to look again at the question of how to make sense of departmentalisation: the question often raised in the past under the banner of the unity of science – or the disunity of science, if you belong to the other party, and under the banner of the autonomy of the special sciences – or the lack thereof, if you belong to the other party.

There are a bewildering variety of answers to our question on the market. Here is a sample. Some say that the account from the physics department – or maybe from the physics department plus that part of the chemistry department that studies physical chemistry – is in principle the whole truth about what our world is like. The suggestion is not of course that the whole truth is given by some claim like 'Electrons exist'. There is a lot more to be said about what our world is like than that electrons exist. Rather the suggestion is that if we assembled everything sayable in terms of physics and physical chemistry into a huge conjunction, we would have the whole truth about what our world is like: our world is an electron at *u* plus a proton at *v* set in a force field of value *w*, *x* cms away from four quarks at *b, c, ,d* and *e,* plus ... – and that's the whole truth about our world. Talk for long enough in the language of physics and do enough conjoining and you will have said all there is to say. Or not quite: you'll need to add at the end 'and that's all'. Or maybe our world is infinite in the sense of being infinitely divisible, or in the sense of being infinite in extent in space-time, or both, in which case no amount of conjoining can do the trick (unless there are patterns in the infinite we can capture with a finite number of summaries). In which case the job of saying it all cannot be done even in principle, but we can say after we've conjoined for long enough, something like 'and more of the same but differently arranged'.

This answer makes urgent the question of what to say about the accounts offered by the other departments: Are they, as some say, mere devices of prediction, to be thought of as heuristics to be used when we lack the real McCoy that physics has to offer; or are these other accounts to be thought of realistically via some program of identifying the posits and concepts they traffic in with the posits and concepts of physics? Others say that we should deny the hegemony pf physics, insisting that each department has its own contribution to make, and that the question of what our world is like is to be answered, not by giving physics some special place, but by conjoining the

accounts from the various departments. Don't reduce biology to physics; set it alongside physics. There are electrons *and* there are cells. Don't reduce politics to physics; set it alongside physics: there are protons *and* there are political parties. Still others say we should reject the whole enterprise, insisting that we are in the grip of what is sometimes called 'one-worldism': the idea that there is one huge, and hugely complex, world that we inhabit, and that we must see the various departments of knowledge as each in their own way seeking to find out what this single world is like. From that metaphysical standpoint, it is inevitable that we find ourselves struggling with the hard question of how to understand the inter-connections between the different accounts coming from the different departments. The way out, according to the deniers of one-worldism, is to turn our backs on that metaphysical standpoint. Each account is fine in its own terms, and that's all we need to say. You can see why this position is sometimes called *quietism*.

We can bring the issues into sharper focus by considering their impact on the debate over what to say about explanations of behaviour in terms of neuroscience, on the one hand, versus explanations of behaviour in terms of psychology, on the other. I might explain the bending of an elbow in terms of the desire for coffee and the belief that bending the elbow is necessary for the ingestion of coffee. But there will also be an explanation of the bending of the elbow in the terms of neuroscience. I cannot of course give you that explanation in any detail but we know that there must be one and broadly what it will look like: it will be in terms of the effect of a certain cup of coffee on a certain brain via the sense of sight and touch, electrical messages from the affected part of the brain to muscles in the arm, and the effect of these messages on the length of certain muscle fibres' and thereby on how much and when the elbow connected to the brain bends.

How are these two explanations of human behaviour – the one in terms of psychology and the one in the terms of neuroscience – related? As above,

there will be four broad classes of response to our question. The first takes the neuroscience explanation to be the real McCoy and treats the validity of the psychological explanation as dependent on its being suitably answerable to the neuroscience one – and here of course the sixty-four dollar question is what 'suitably answerable' comes to, but it will be some sort of reduction of the story as told by psychology to the story as told by neuroscience. The second sees the explanation in terms of psychology in instrumentalist terms. Positing belief and desire allows the prediction of behaviour – something one could hardly quarrel with – but insists that they, the beliefs and the desires, are merely convenient fictions that make prediction easier; it is a confusion to ask if 'they' really exist. The third is the conjunctive response: both explanations are fine in their own terms and we should rejoice in the fact that we have two distinct explanations of the same phenomena. The fourth response insists that we are asking a bad question. There is no single way things are as far as the causation of behaviour goes that demands we reconcile somehow or other the two apparently different explanations of behaviour.

In this lecture I will tell you about the answer I favour to the departments of knowledge question. This is one of the things that philosophers can do. The physics department can tell you very much more that I can about physics – and a corresponding remark applies for all the departments I mentioned. The contribution a philosopher can make is to the meta-question of how to make sense of the inter-relations between the accounts that come from the different departments, thus the title of this lecture.

The case for one-worldism and against many-world views

The answer I favour takes off from what we called above 'one-worldism', the view that we inhabit a single huge world. It is part and parcel of an idea most especially associated with W. V. Quine: although many different kinds of things exist, this is a difference in kinds of things, not in kind of existence. The

difference between tables existing and nations existing lies in the tables and the nations alone. This position is current orthodoxy in analytical philosophy and many insist that it is about as obviously true as anything can be. I in fact find myself in this situation. But now I have a problem. How can I *argue* for it? There is nothing more secure from which to argue, as I see things.

What I can do is argue that many-world views do not solve the problems we need solutions for in this area. They don't buy us the answers we need. I will make the point with some examples.

Sometimes one causal explanation excludes another causal explanation; sometimes it doesn't. An example where exclusion happens is the way vitalism in biology was overthrown. The discovery that it was cell division that explained the growth of plants was rightly interpreted as showing that the view that 'vital forces' did the job was a mistake. More generally, once we discovered that the distinctive behaviours of living things could be explained without recourse to a special vital force, we deleted vital forces from our ontology. Similarly, the explanation of the operation of pumps in terms of air pressure excluded explanations in terms of abhorrence of a vacuum, and oxidation excluded phlogiston as an explanation of combustion. But in other cases, exclusion doesn't happen. To take a famous example, the discovery that certain molecular properties explain the behaviour of gases was not taken to exclude explanations in terms of temperature and pressure. Instead, temperature and pressure were identified with the relevant molecular gas properties. Identification happened, not exclusion. The molecular kinetic theory of gases did not displace the properties of the thermodynamic theory; it instead told us what they were. Likewise explanations of the way in which elements combine in terms of valency survived explanations in terms of atomic structure. Instead of displacement we got illumination.

It is easy to see on many-world views how *failures* of exclusion can happen.

Why should what is said by a department addressing the nature of the explanation obtaining in one world adversely affect what is said by a department addressing the nature of the explanation obtaining in another world? The puzzle for many-world views is rather to explain why exclusion ever happens. The same problem arises for certain incommensurability views. If what is said by one department is *incommensurable* with what is said by another, how come sometimes what is said by one excludes what is said by another? And if incommensurability reigns, how could we possibly address in a principled way when exclusion happens and when it fails to happen? The more one talks about incommensurability, different frameworks, different worlds, autonomous conceptual schemes that only answer to their own constraints, and so on and so forth, the harder one makes the task of understanding how it is that we ever get cases where an explanation in terms of one set of concepts excludes one in terms of a different set of concepts, and the harder one makes the task of giving principled accounts of the difference between the cases where we get exclusion and cases where we don't.

The case against naive one-world views with conjunction

Interestingly, the same cases make trouble for naive versions of one-world views with conjunction. By views of this kind I mean ones that hold that we should accept what the different departments say about what our world is like and simply conjoin. It is naive in that we *simply* conjoin. It is, if you like, another form of *quietism*. The trouble for naive one-worldism is that the cases we have recently discussed tell us that often we should not conjoin. The account offered by one department excludes the account offered by another. We should not conjoin the account of plant growth offered by vitalism and that in terms of cell division offered by modern biology. We should take the second to falsify the first. But let me focus the discussion a bit more by making the point in terms of a variant on our earlier example of the relation between explanations of subjects' behaviour offered by psychology and those

in terms of the internal workings of subjects.

There is information in physical terms about the internal workings of things that look exactly like us, and move through the environment much as we do, that would reveal that they lack a psychology despite the fact that we would find it very natural to credit them with mental states. Humanoids controlled from Mars, as in a variant on Christopher Peacocke's Martian Marionette case, and creatures that work by look-up tree, as in Ned Block's famous example, lack a psychology precisely because they work in the wrong way for having a psychology, no matter how good a job they do of passing as us 'from the outside'.[1] 'Explaining' their behaviour in terms of psychological states would be instrumentally justified and very natural but, all the same, would be quite wrong; they have not a thought in their heads. But of course it would be a mistake to infer from this that anything for which there is a sufficiently complete account of its internal workings, and how they lead to behaviour, giveable in physical terms, lacks a psychology. We have a rich psychology, and yet, as we noted earlier, we know that there is such an account of our internal workings and their relation to the environment that would explain our behaviour.

---

[1] Christopher Peacocke, *Sense and Content*, Oxford: Oxford University Press, 1983, and Ned Block, 'Psychologism and Behaviorism', *Philosophical Review,* 90 (1981), pp. 5–43. A creature that works by look-up tree is like someone who doesn't understand what a square root is but gets the answer right whenever asked for a square root by looking up the answer in a table of square roots, with the difference that a creature that works by look-up tree does this for any and every challenge posed by their environment. They respond as instructed by the entry against any given environmental challenge in a *huge* look-up tree or table. The case is possible in the sense that we can understand it; it is not possible in any stronger sense, as the needed table would involve a combinatorial explosion.

The upshot is that we cannot go quietist. Sometimes what is said by the department of neuroscience contradicts what is said by the department of psychology – that happens when neuroscience reveals that a subject is controlled from Mars or works by look-up tree – and sometimes it doesn't; and we need to know what makes the difference between the two kinds of case. We need, as I will say it, to be sophisticated one-worldists and not naive one-worldists.

Sophisticated versions of one-worldism: emergence, modest and immodest

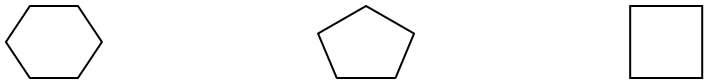
The rest of this lecture is concerned to give the version I favour of how to understand the inter-relations between the deliverances of the various departments of knowledge, of how to be a sophisticated one-worlder. It will involve sketching for you an overarching perspective on the debate over the special sciences.

We start with a distinction between two sorts of emergence.

By emergence I mean the way aggregation generates new properties in the sense of new patterns in nature. Aggregations have properties that outrun those of their parts and the ways of aggregating those parts. A triangle is an aggregation of items that lack triangularity. A house is an aggregation of items none of which has the property of being a house. A powerful car is an aggregation of items none of which has the property of being powerful.

The distinction between modest and immodest emergence can be illustrated by reference to the kind of exercise that is common in construction problems in geometry.

In these problems we are given a limited stock of basic materials and a limited range of ways of operating with those materials, and are asked what kinds of things can and cannot be constructed. For example, we know that with a ruler and compass alone, it is possible to construct a regular polygon of n sides for n = 3, 4, 5 and 6, but not for n = 7. We also know, famously, that it is not possible to square the circle. However, to fix our ideas we will take a very easy example. Given an unlimited supply of finite straight lines of any length, all lying in a plane, what closed figures can be constructed simply by moving the lines around in the plane? The answer is that we can construct figures like



but we cannot construct figures like

.

Now in construction problems in geometry the following is true: the new kinds or patterns that can be constructed are *a priori* determined by the nature of that which is aggregated and the ways of aggregating. It is, for example, an *a priori* matter what can and what cannot be made by re-arranging finite line segments in a given plane. *Modest emergence theses* hold that any new properties that come from aggregation are *a priori* determined by what is aggregated and how it is aggregated. *Immodest emergence theses* hold that at least some of the new properties that come from aggregation outrun those that are *a priori* determined by what is aggregated and how it is aggregated. I think we should believe in modest emergence rather than immodest emergence. Or as I say it elsewhere, I think we should accept a certain closure

principle for properties generated by aggregation: they are closed under *a priori* determination.[2] I would love to have an apodictic proof of this closure principle but lacking this, I offer three considerations.

The first is Occamist. Ontological austerity favours minimising the number of new patterns one should acknowledge as coming from aggregation. We have no choice but to accept that which comes *a priori*; indeed, we can even minimise the ontological cost by following David Armstrong and declaring the *a priori* an ontological free lunch.

The second consideration is that I know of no compelling example of an emerging property that is not *a priori* determined by that on which it emergences. Every example where aggregation generates new property instances is one where the new property is *a priori* determined by what is aggregated and how it is aggregated. Let me illustrate the point with my favourite example. Despite their complexity, we know that the dances of honey bees fall into various patterns or kinds. If that were not the case, the dances would fail in their evolutionary purpose of providing the bees with useful information about the location of honey. When Karl von Frisch and others set out to decode the dances, they broke the dances and their environmental settings into arranged components in their search for the patterns. And they rightly took it for granted that finding the answer was a matter of finding the properties of the ways the bits go together that *a priori* determine the patterns in question. The possibility of having a *sui generis* pattern that emerges without being *a priori* determined by the nature of the aggregated elements was given no credence whatever.

---

[2] Frank Jackson, ' On Ensuring that Physicalism is not a Dual Attribute Theory in Sheep's Clothing', *Philosophical Studies*, forthcoming 2006.

The third consideration is epistemological. I cannot see how one could have any good reason to believe in the emergence of new properties over and above those that are determined *a priori*. Let me illustrate the point with a live debate among cognitivists in ethics over the relation between moral and non-moral properties.

The supervenience of the moral on the non-moral is the compelling contention that if *x* and *y* are exactly alike in non-moral respects, they are exactly alike in moral respects also. Equivalently, they cannot vary in moral ways without varying in non-moral ways. It follows from the supervenience of the moral on the non-moral that every moral property is necessarily co-extensive with a disjunction of non-moral properties. Here's one way to see this. Supervenience tells us that the non-moral determines without remainder the moral – if it didn't, we could vary the moral independently of the non-moral, and supervenience says we can't. Now consider every right act in logical space and suppose that 'N_1', 'N_2', ... give the non-moral nature of each act, respectively, then the following bi-conditional will be necessarily true

x is right iff x is N_1 v N_2 v ...

There are four possibilities for the infinite disjunction on the RHS. One, there is no pattern to the non-moral disjuncts: that is tantamount to going eliminativist about ethical properties. Two, there is a unifier but it is an extra property of the kind G. E. Moore believed in. That is a metaphysical implausibility. Three, there is a unifier; it is not an extra property and it is *a priori* determined by each disjunct. This in fact is the view in ethics I favour. Finally, there is a unifier; it is not an extra property but it is *not a priori* determined by each disjunct. This last is the position of the anti-reductionist naturalists in ethics: no metaphysically objectionable non-natural properties to be the moral properties but no *a priori* connection between the moral and non-moral properties either. My problem is with the epistemology of such a position. How could we ever be justified in believing in the existence of this

unifier? Supporters of anti-reductionist cognitivism in ethics talk of 'seeing' the patterns in the non-moral that are the moral properties through the eyes of a morally informed perspective on the world: what's invisible from the non-moral stance becomes visible from the moral stance. To me that sounds awfully like claiming to see extra properties, and that, as anti-reductionists themselves acknowledge, is a very dubious bit of metaphysics.

We can now state the version of sophisticated one-worldism I favour. It runs as follows. There is one huge, and hugely complex, world that we inhabit and there are two sources of patterning in this hugely complex world. One source lies in the basic ingredients. There is a finite number of kinds of fundamental ingredients. The fundamental ingredients are five in number, or twenty-five or .... The exact number isn't important and in any case is not something I am any kind of expert about. What is important is that there is a finite number. The fundamental ingredients are the items whose aggregation makes up our world in all its variety. Our world is a vast aggregation or merelogical sum of those ingredients, where the ingredients include modes of aggregation as well as the items aggregated. Of the departments we mentioned right at the beginning, it is physics and physical chemistry that are most likely to be the ones that tell us about these fundamental ingredients.

Physicalism about the mind and consciousness is sometimes put by saying that it is the doctrine that God could make minds and consciousness by suitably assembling purely physical ingredients alone. He or She would not need to add consciousness or sentience. They are there automatically if the physical ingredients are assembled aright, just as a car is powerful if it is the right kind of aggregation of parts. There is no need at the end of the assembly line to have a stage where the 'powerfulness' gets added. Following this lead we can put what I have just said about the nature of the world we live in thus: God could make our world down to the last detail simply by suitably assembling the fundamental ingredients aright; there would be no more for

Him or Her to do.

The second source of patterning in our hugely complex world lies in the way aggregation generates new properties, in the way new properties emerge. They one and all emerge by being *a priori* determined by the nature of that which is aggregated. We have, that is, two constraints on complexity: a constraint on the components available to generate new properties – there is a finite list; and a generation constraint – new properties are *a priori* determined by the relevant aggregations.

We now have a simple way of looking at the many controversies over the departments of knowledge – or in the more usual way of putting matters – over the status of the special sciences.

The super department and the special departments

The super department inventories the fundamental ingredients. If we ever had the total account of what our world is like in these terms, we would have the complete account of what our world is like, in the sense both of knowing that which determines without remainder what our world is like, and in the sense of knowing that which *a priori* determines what it is like. In this sense the picture is reductionist but, as we will see very shortly, it is strongly anti-reductionist in other important senses.

The special departments are concerned with the properties that come from aggregation, the properties which emerge. Aggregate one way and you get nations and your department is politics; aggregate another way and you get cells and your department is biology; and so on. Their laws are the laws of the emerging patterns and, as properties really do emerge and as these emerging properties really do fall under laws, our picture is in that sense strongly anti-reductionist. The laws of the special sciences really are laws. The easiest way

to see this is with a very simple example.

Suppose, contrary to fact, that the super department is the department of Newtonian physics applied to point masses, and consider the evolution of a system of isolated particles over time. We know that the position, velocity and mass of each particle at any point in time fully determines the position, velocity and mass of each particle at any subsequent time, in accord with Newton's laws of motion. We also know something about the total momentum of the system: it remains constant. Now the total momentum of the system is not one of the fundamental properties. It is an emerging property; it is a property of an aggregation, namely, the system of particles. Our very simple example thus tells us that emerging properties are perfectly real – total momentum is a real property of systems of particles, and that emerging properties can fall under real laws – conservation of total momentum is a perfectly good law in Newtonian physics despite not being a law that governs the properties of the fundamental ingredients. It follows that nothing in our picture impugns the reality of the properties of the special departments and nothing in our picture impugns the lawfulness of the laws of the special departments.

Our picture also supports an important epistemological point about the knowledge provided by the special departments. Although emerging properties are fully determined by the properties on which they emerge, it does not follow that the best way to find out about emerging properties goes via the properties they emerge on; nor does it follow that the best way of finding the laws governing the emerging properties goes via the laws governing the properties they emerge on. Indeed, often that would be a very bad way of finding out about the emerging properties and laws. Metaphysical priority and epistemological priority come apart in many cases. The reason lies in the complexity of the underlying properties: often what emerges is much simpler than its emergence base. Our example of the conservation of

momentum makes the point; so does Hilary Putnam's famous square peg-round hole example.[3]

When a square peg fails on a number of occasions to go through a round hole whose diameter equals the side of the peg, there will, on each occasion, be an explanation in terms of the positions of the constituent atoms and the forces that tie them together, and more generally the micro-structural features, of the peg and the hole. This will be a perfectly good explanation of why the peg fails to go through the hole but it has two serious disadvantages from an epistemological point of view. First, it is very unlikely that anyone will know the precise positions of the constituent atoms. Nearly always there will exist an explanation of such and such kind that is available in principle, but not in practice. Secondly, the explanation makes opaque an important modal property of the situation. There are many ways of trying to pass a square peg of side x through a hole of diameter x, and each failure will have a very different explanation in terms of the positions of the atoms and the like. For example, the particular atoms that do the 'blocking' will differ greatly depending on the angle of approach. However it would be a bad mistake to think that there was no 'uniter' for the failures: there is but it lies in the emerging properties. It lies in the shape of the peg and the hole; that is, in the properties of certain aggregations of atoms. I am not saying there is no uniter at the micro-level (though some do seem to say this about the example). What unites the phenomena at the micro-level is how far apart various atoms are from one another, the rigidity of the lattices they compose and the fact that they make up square and round arrays.[4] What we learn from the example –

---

[3] Hilary Putnam, 'Reductionism and the Nature of Psychology', *Cognition*, 2 (1973): 131–146.

[4] More precisely, the relevant subvenient pattern in the peg-hole case is: for every line through the peg, the set of cross-sections orthogonal to that line

and the momentum example – is that explanations in the special sciences can have *enormous* epistemological and methodological advantages over explanations in the terms of the super department.

Our approach also allows us to give a simple consistency proof for the special sciences. The various special sciences trespass on each other's territory. They offer different explanations for the very same facts. That is one reason, as we saw, for rejecting many-worldism and incommensurability about the special sciences, and this is the point Daniel Dennett made vivid for us with his distinction between the physical stance and the intentional stance in explanations of the very same phenomena in psychology.[5] His recipe for stopping us worrying about this situation was to go instrumentalist about the intentional stance. The physical stance is the 'realistic' description and the reason there is no clash with the apparently disparate explanations of, as it might be, my being here in Hong Kong, given by the physical and intentional stances, is that the intentional stance is not a competitor that posits different causes from those posited in the physical explanation, but a mere predictive device.

Our approach has the advantage of allowing a consistency proof without going implausibly instrumentalist. We borrow from consistency proofs in the formal sciences. To prove that a theory T_1 is consistent with a theory T_2, we find a model that makes both true. On our approach we get this for free. The 'model' is the account of what our world is like given by the super-department. We said earlier that, out of our current crop of departments, physics and physical chemistry have the best chances of being the progenitors of such a super-department. For ease of exposition, let's christen the super-

---

contains at least one with rigid points that are further apart than the diameter of the hole.

[5] Daniel Dennett, *The Intentional Stance*, Cambridge, Mass: MIT Press, 1987.

department that gives an account of our world that covers everything in the sense that every and anything that is to be found in our world is nothing over and above an aggregation of the items in that account, 'physics' and its items 'physical'. So any and everything in our world is an aggregation of physical items and has the properties that emerge on this aggregation, and as I've argued, these properties are closed under *a priori* determination. The special sciences are then the accounts of our world giveable in terms of the various relations that obtain between the emerging properties. Different special sciences differ in the levels of aggregation. Aggregate one way and you get one set of emerging properties that stand in one set of relations; aggregate in another, and you get another set of properties (typically not a disjoint set) that stand in another set of relations (again typically not a disjoint set). And so on. Thus, each special science is made true when it is true by the model that consists of the vast aggregation of each and every physical item. We have our consistency proof.

What happened when cell biology displaced vitalism as an account of the behaviour of living things? We discovered that the model that makes cell biology true makes vitalism false. Why did the kinetic theory of gases fail to displace the thermodynamic theory of gases? We discovered that the model that makes the kinetic theory true makes the thermodynamic theory true also.

The famous autonomy issue

My interest in the relation between the knowledge we get from the various special departments of knowledge was originally prompted by the heat I discerned in debates over the autonomy of the special sciences. Part of what I have been doing in this lecture is a 'divide up the issues and smooth the waters' exercise. If you mean X by autonomy, the answer is pretty obviously so and so, whereas if you mean Y, the answer it is pretty obviously such and such. So shouldn't we all agree? For example, who could possibly disagree

about the epistemological advantages of explaining the failure of the square peg to go through the round hole in terms of the geometry of the shapes, over the explanation in terms of the various locations of and connections between the individual atoms that make up the peg and the hole's surrounds? And who could possibly disagree about the possibility of special science laws that are laws in the fullest sense – the example of conservation of momentum shows that?

What then is there to fight over? As far as I can see just two issues. One we have discussed at a little length: the principle of *a priori* kind closure for emergent properties. I won't say anything more about that. The other issue is the question of the extent to which some one or another special department, *D*, is answerable, in the last resort, to what is said by the super department, and, in the medium term, to what is said by departments concerned with the behaviour of the items which compose or make up the items *D* is concerned with. We touched on this issue when we discussed creatures that work by look-up tree and creatures controlled by messages from Mars. They were both examples where psychological explanations of one or another creature's behaviour were trumped by information about the behaviour of certain bits that are parts of the creatures in question's composition. I will close by making some quick general remarks about this question.

Consider again the square peg-round hole example. We saw that the explanation in terms of the respective shapes of the peg and the hole, and the explanation in terms of the locations and relations between the constituent atoms were both good explanations, with the first having two signal advantages: we know the shapes but do not know the locations of the various constituent atoms, and the first explanation reminds us that it would not have made any difference had we tried to get that peg through that hole along some other line of approach. Despite these important advantages, there is an important sense in which the explanation in terms of shapes is answerable to

the explanation in terms of atoms and their locations and relations. Suppose that when we dig into the causal mechanics of the peg's failure to go through the hole, it turns out that what is happening is that atoms of this kind when clustered together generate a force field just in front of the peg, and it is this force field that stops the peg going through the hole. What we learn, that is, is that the shapes are irrelevant. What does the causing is the force field. The shapes *would* have been relevant had there been no force field but they are in fact irrelevant.

Or consider explanations of the inheritance of characteristics in terms of recessive and dominant genes. No matter how well-founded these explanations are, they are answerable to the mechanics; they are answerable, that is, to the details of how various parts of the bodies of those who pass on their characteristics causally interact. And of course some evolutionary biologists take an instrumentalist approach to genes, regarding the recessive-dominant gene story is a convenient fiction and not the literal truth of the matter about the passing on the characteristics from one generation to the next; but they do this precisely because they think that discoveries about the mechanics of the process show that genes do not exist. I don't agree with their conclusion – I think they get their result by setting the standards for being a gene unreasonably high – but I do think that they are right that the gene story is answerable to the details of the interactions of the parts.

So in this last sense, the lender of last resort sense if you like, the special sciences are not autonomous. What makes their explanations correct when they are correct lies in the facts as captured in the terms of the account of the superdepartment, the department we have been calling physics.