Chapter 2 Reduction and Selection

2 .1 Introduction

In the last chapter I upheld the orthodox view that reducibility to physics, in the sense of type identity, is too strong a requirement for the categories of most special sciences, and for those of psychology i n particular. In this chapter, however, I want to show that the case for reducibility to physics is rather stronger than is generally recognized. More specifically, I want to show that this case is compelling for those special sciences that do not have a teleological underpinning. As it happens, I think that psychology in particular does have such a teleological underpinning, and that its categories are therefore not reducible to physical categories. But my argument will imply that special sciences without a teleological underpinning are indeed reducible to physics.

2.2 An Unexplained Coincidence

It will be convenient to begin with the functionalist view of mental states. As I explained in the last chapter, functionalists view mental states as causal intermediaries between perceptual inputs and behavioral outputs. According to functionalistm, you will be in a given mental state as long as you are in a physical[1] state which plays the relevant causal role between perception and behaviour.

It is widely regarded as a great merit in functionalism that it leaves room for irreducibility, and allows that mental states should have different physical rea lizations in different people, or even in the same person at different times. According to functionalism, what is common between between John Major and Boris Yeltsin, when they each believe that there is an ice-cream in front of them, say, is that t hey are each in some physical state which is characterictically caused by the presence of an ice-cream, and which characteristically causes them to reach out if they want an ice-cream. But functionalism doesn't require that this be the same physical state in both cases -- which is just as well, given how unlikely it is that there should be some strictly physical feature common to all and only those people who believe that there is an ice-cream in front of them.

However, there is som ething rather puzzling about the picture that functionalism now invites us to accept. If states like believing there is an ice-cream in front of you, and wanting that ice-cream, are realized by different physical states in different people, then why do these states always have the same behavioural effect in all those different people, namely, reaching out for the ice-cream? In general, we expect physically similar states to have similar effects, and physically different states to have differen t effects. So why in this case should physically different states have the same effect?

Consider an analogy. Imagine that people forced to eat a certain restricted diet -- nothing but reheated brussels sprouts, say -- invariab ly develop certain characteristic symptoms -- inflamed ankles and knees, say. Nutritionists investigate this phenomenon. But they find no uniform explanation. In one case, the sprouts harbour a virus which flourishes in the ankles and kn ees and provokes the immune system. In

another case, eating the sprouts leads to excess production of uric acid and hence to gouty attacks. In another, the diet leads to a nutritional deficiency which depletes the cartilage which protects the joints. And so on. For each person, there is some physiological explanation of why the diet leads to the inflammation, but the explanation is different in each case. I take it that this would be incredible. If the diet triggered ei ther of just two different sequences, say, both of which then happened to cause inflammation in the ankles and knees, we could perhaps view this as a curious coincidence, an amusing oddity with no further explanation. But that it should trigger an i ndefinite number of quite different sequences, yet all of them lead to the same inflammation, would surely be quite absurd, in the absence of further explanation.

Yet the thesis of variable realizability seems to commit us to something quite analogous , namely, that the same perceptual inputs give rise to quite different internal states in different people, and yet those different internal states will all end up generating the same behavioural outputs. This too is surely quite absurd, in the abse nce of further explanation.

Contrast the functionalist[2] picture with the kind of situation where physical reduction is possible, as when kinetic theory reduces the classical gas laws to the basic dynamics of molecular movement. At fir st sight it mightn't be clear how this kind of case differs from the functionalist picture. After all, aren't there lots of different ways in which the molecules can be moving around in a gas at a given temperature, thus giving us a heterogeneity of physical states for the single macro-state of having that temperature?

But, even so, there is still something physically in common between all those different physical states, namely, that the molecules have a given mean kinetic energy. It is t his commonality that then enables us to explain such things as why an increase in temperature at constant volume always results in an increase in pressure.

Reducibility to physics does not involve the absurdly strong requirement that the instances of the reduced category should share all their physical properties. The requirement is only that there should be some physical property present in all and only those instances, which then allows a uniform physical explanation of why those instances alw ays give rise to a certain sort of result.

But that is precisely what we don't have in the functionalist case. If there is nothing physically in common among the realizations of a given mental state, then there is no possibility of any uniform e xplanation of why they all give rise to a common physical result. And that's what I find puzzling.

Imagine that the temperatures and pressures of gases were always realized by internal molecular motions, and temperature increases always led to p ressure increases, but yet it was impossible to explain this in terms of basic physics. I take it that this would be incredible. But that's what functionalism is asking us to believe about psychology.

It is worth emphasizing that I am not accusing the functionalist picture of inconsistency, but only of incredibility. The dificulty I am concerned with arises when some mental state S, which mediates between physical input R and physical output T, is realized by a range of different phy sical states Pi. The puzzle is: why do all the different Pis which result from R all nevertheless yield the common effect T? Now, it is possible that every such Pi should just happen to yield T, just as it is

possible that all the different ph ysical consequences of eating reheated brussels sprouts should just happen to cause inflamed ankles and knees. However, if this were so, it would be the kind of coincidence that cries out for explanation.

At bottom, the difficulty I am raising i s an empirical one. Our experience of the world has shown us that if a certain physical result always appears after certain physically specified antecedents, then there is always some uniform explanation in terms of physical laws. But the func tionalist picture violates this general principle. It commits us to the existence of a physical generalization, namely, that R leads to T, but denies that it can be explained in physical terms. I think this ought to make us think again about f unctionalism. (For other versions of this argument, see Papineau, 1985; Searle, 1985, ch 5; MacDonald, 1986, sect II.2.)

2.3 Laws in the Special Sciences

Although the last section focused on functionalism i n the philosophy of mind, the problem at issue is clearly generalizable to any category in any special science which (a) is related by empirical law to physical antecedents and physical consequents yet (b) is variably realized at the physical level. For in any such case we face the same puzzle of why all the different physical realizations of the special category should give rise to the same physical result.

One obvious way of resolving this puzzle would be to deny (a), that is, to deny that the special category in question is related by law to physical antecendents and consequents. For there obviously won't be any puzzle about how the different physical realizations of some special S all produce the same physical result in appropriate cir cumstances, if there isn't any such result that they all produce.

It do not propose to adjudicate how far this move is plausible for the different special sciences. The question of the existence of psychological, social and biological laws is a standard topic in the philosophy of these special sciences, and there is no question of engaging with the huge literature on these questions here. But what I shall show in this section is that, if you do want to resist reductionism by denying the existence of laws, then your denial needs to be whole-hearted. It is not enough merely to maintain that the laws of biology, say, are less strict than those in basic physics. For even lax biological laws will be puzzling, if there are no r eductive relations between biological categories and physical ones.

Let me illustrate this point by considering the position adopted by Jerry Fodor in his infuential article "Special Sciences" (1974). In that article, Fodor defends the general f unctionalist picture I have been concerned with so far: he takes it that any special S will be realized on different occasions by different physical Pis, but that nevertheless such special Ss can enter into laws linking them with subsequent resultsR , in virtue of the fact that processes operating on the physical level will generally lead from each Pi to R.

However, Fodor adds a twist, which might seem to avoid the difficulty I have raised for functionalism: Fodor insists that the law linki ng S with result R will have exceptions. On Fodor's picture, not all the Pis which realize S will give rise to result R, and in consequence the S-R law will not be invariable. So Fodor seems to have an

immediate answer to the question of why a ll the different realizations of S yield the common result R -- they don't.

However, note that Fodor continues to hold that S usually leads to R, or tends to lead to R, or some such. And this in itself raises a puzzle, in the absence of any conc essions to the reducibility of S. For if there is no common physical pattern at all to the realizations of S, then it will be puzzling that there is even a tendency for S to lead to R.

By denying that laws involving special categories are exact, Fodor can resist the argument for an exact reduction. But he still faces an argument for an approximate reduction, as long as his special science contains approximate laws. For even such approximate laws will be puzzling, unless there i s some common physical category which usually realizes S, and thereby explains why S is usually followed by R.

There is, I think, a general pattern here. By weakening the extent to which a special science contains general truths, you can weaken the extent to which its categories have to be reducible to physics. But you can only avoid reductionist conclusions completely by denying that the special categories enter into general lawlike patterns at all.

By way of further illustration, con sider a suggestion made by Davidson. In general Davidson is sceptical as to whether any serious laws can be framed using psychological terminology. However, in "Hempel on Explaining Action" (1976) he offers the suggestion that the generalizati ons involved in explaining and predicting actions are always person-specific. We can know that Jim, say, will buy an ice-cream when he wants one, will lose his temper if he thinks someone has been rude to him, and so on. But at the same time t here are other people of whom these things aren't true. And so, even if we can know a law which applies to Jim, this doesn't mean that there are psychological laws ranging over people in general.

My immediate concern here is not to evaluate this ingenious suggestion, but just to point out that the reductionist argument still gets some grip on even this minimalist Davidsonian conception of psychology as a generalizing science. Davidson still holds that it is a general truth that if Jim want s an ice-cream, he buys one. And this itself would be mysterious unless there is at least a uniform physical realization of Jim wanting an ice-cream. (In fairness to Davidson, it should be noted that he takes the relevant generalizations to be dispositional as well as person-specific. This raises further issues which I discuss in the next section.)

It is true that a uniform physical realization for Jim wanting an ice-cream doesn't amount to a very strong form of reduction. But it's still something, given that even Jim wanting an ice-cream can in principle be realized by different physical states in different instances. Once more, the moral is that, insofar as generalizations of any kind are admitted in a special science, to that extent there will be an argument for a corresponding amount of reduction.[3]

2.4 Functionalism and Dispositions

There is one obvious way in which S could be variably realized and yett here be no mystery about a general law linking S and R. Namely, if S were a dispositional term,

defined as the state of being disposed to give rise to R in appropriate circumstances. In that case there wouldn't be any further need to explain, via some physical reduction of S, why the different realizations of S all give rise to R: giving rise to R is precisely what makes those different states all count as realizations of S in the first place.

It is possible that this thought has ten ded to obscure the difficulty that I have been raising for functionalism. Functionalism makes it a matter of definition that any given mental state gives rise to specified behavioural effects. And so, if you focus on this aspect of functionali sm, it may seem natural to conclude that there can't be any real difficulty about how such states can be differently realized physically, and yet have common physical effects. Isn't this just an upshot of their functionalist definitions?

However my puzzle can't be dissolved that easily. The basic difficulty is that functionalist concepts aren't so much dispositional concepts as theoretical ones. Functionalism defines special categories not just as states which produce certain effects, but rather as states which enter into a certain structures of causes and effects. According to the functionalist picture of psychology, for example, pain will be defined not just as the state that characterisically causes avoidance behaviour, but also a s the state that characteristically results from bodily damage; again, the belief that there is an ice-cream in front of me will be defined not just as the state that characteristically causes me to reach out if I want an ice-cream, but is also the state that characteristically results from my looking at an ice-cream with my eyes open. In general, functionalist definitions of Ss allude to their resulting from Rs, as well as to their causing Ts.

This means that the argument against S being variably realized now goes through as before. We can put it like this: if the various realizations of the state which arise from R have nothing physically in common, then how come they all alike give rise to T? If the various realization s of the state which arises when people look at an ice-cream have nothing in common, how come they all alike lead to their reaching out for it?

The presence of R as an independent criterion for the presence of an S, independent of S's effects, means t hat we can't any longer simply account for S's realizations all yielding T by saying that's what makes them realizations of S. For now something else also makes them realizations of S, namely, that they arise from R.

It is worth emphasizing that the position which I am claiming faces a difficulty is not functionalism understood merely as the claim that psychological states can be defined in terms of structures of causes and effects. Rather, the difficulty arises specifically when functiona lism in this sense is combined with the thesis that psychological states are variably realized. So I have no argument against the philosopher who insists that our concept of pain is a concept of a second-order state defined in terms of certain perce ptual causes and behavioural effects, and not a concept of any specific physical state. My concern is only to point out that, unless there is a specific physical state which generally realizes pain, albeit an unknown one, it would be a puzzle why th ose perceptual causes are generally followed by those behavioural effects.

A connected point. I am exploring an argument for the conclusion that the categories of the special sciences must be reducible to physics. This is not an argument, however, for the conclusion that the practitioners of the special sciences have to know

that reduction. For, even if such a reduction is in principle available, you don't necessarily have to know it to have an adequate conceptual grasp of the releva nt special categories, and hence be in a position empirically to investigate them. After all, the classical gas laws were well known long before kinetic theory was developed. (This point would be too obvious to be worth making, were it not for my suspicion that many people are attracted to "functionalism", in the strong sense of variable realizability, because they think that without it the special sciences would be under an unfortunate immediate obligation to produce physical reductions of th eir categories. But of course there are plenty of other possible justifications for not producing immediate reductions, apart from functionalism in this strong sense.)

David Lewis (1980) combines a functionalist definition of psychological conce pts with the view that those states are uniformly realized in any given species. Up to a point this position circumvents the difficulty I am raising. Within any given species, so to speak, there is no puzzle as to why the state that arises fro m R always gives rise to T, for by Lewis's own account that state will be a homogeneous physical state which will lead to T as a matter of physical law. However, there does remain a problem across species. If the central states that result fro m bodily damage in octupuses and and frogs and humans are all so different, how come they all lead to movement away from the external cause of the pain? Of course, behavioural or neurophysiological observation of each such species could show us that the various central states in question all give rise to such avoidance behaviour, that is, could show us that all these species had states that fitted the general functionalist definition of pain. But, without further explanation, there would still be a puzzle as to why, despite their physical differences, the different central states that arise from bodily damage should have the same physical effects. (By now a solution to this puzzle will no doubt be suggesting itself. Namely, that th ese states all have the same effect because they have all been naturally selected to produce that effect. But let us leave this solution until section 2.7. My current concern is only to establish that there is a puzzle here to be solved.)

2.5 The Irreducibility of Ordinary Dispositions

The argument of the last section showed that there is no reason to suppose that dispositional concepts in general will be physically reducible. Provided a given d ispositional concept doesn't enter into any further laws, in addition to the definitional "law" that connects it to its display, then we have no reason to expect reduction. For example, redness is arguably definable as a dispositional characteristic of objects, namely, the characteristic of producing a certain kind of perceptual response in normal observers. But it remains perfectly possible that there is nothing physically in common between all the different objects that produce this response . If the notion of redness entered into certain kinds of further laws, then there would be reason to expect a reduction. But the mere fact that all red things make normal people see red doesn't itself give reason to expect reducibility.[4]

Again, the biological notion of fitness is arguably definable as a dispositional characteristic of biological traits, namely, the characteristic of enhancing survival better than alternative traits influenced by the same genetic locus. But it would be wrong to expect that just because of this there will be anything physically in common between different fit traits and the ways they enhance survival.

Of course, if we want to explain the display of a disposition by reference to that disp osition itself, as when we explain someone's visually judging something to be red by reference to its redness, or when we explain some trait's selection by reference to its fitness, then we will be committed to the disposition as something more than just the property of producing that effect. But this commitment to "something more" could just be that there is some physical basis for the production of the effect. And this commitment can thus leave it open that the physical basis might be differ ent in different instances of the dispositional property.

Let me emphasize the requirements for the irreducibility of a dispositional property. Such irreducibility requires that the property not enter into any substantial non-dispositional laws, that is, that there not be any uniform physical cause for the different physical bases of the disposition, nor any uniform physical effects which aren't themselves effects of the purely dispositional definition (for in this latter case we would the facet he non-definitional issue of why all the different realizations which are definitionally grouped together by their dispositional display also always have another common physical effect). These are fairly strong requirements, but I see no reason to s uppose that they are not satisfied in plenty of familiar cases. For example, I take it that redness does not in fact have any such uniform causes or uniform effects. There is surely no uniform physical cause for all the different physical base s for redness, nor, arguably, any uniform effect of redness independently of its uniform effect on observers. And similarly with fitness: there is no single cause of all the different physical properties which make different traits fit, nor an y uniform effect of those properties apart from their influence on survival. Which is why, once more, there is no reason to expect redness or fitness to be reducible.

We might wonder why dispositional concepts are useful, if they cove r a heterogeneity of different physical bases, and have no uniform causes or effects apart from their defining display. However, there are obvious reasons why it might sometimes be useful to classify things together just in virtue of their producing a certain kind of effect. An interior designer may not care about the molecular constitution of a fabric, nor even about how it was made, but simply about the colour responses it will produce in humans. Again, suppose we are interested in pre dicting the spread or extinction of some biological trait. All we need to know is its fitness, not its physical nature or its developmental history. If two different traits have the same fitness relative to their competitors, then they will ev olve in the same way, whatever the other differences between them.

2.6 The Meaning of Reduction

One question sometimes raised in the literature is whether there is really a principled difference between reduction and variable realization: for, in a case of variable realization, why shouldn't we simply disjoin the various Pis which realize S, and then say that S reduces to (P1 v P2 v . . . v Pn)?

In "Special Sciences" Fodor responds to this challenge by saying that the disjunctive property (P1 v . . . Pn) won't in general be a genuine natural physical kind, and the generalization (P1 v . . . Pn) -> Q won't therfore be a genuine law. However, Fodor's

analysis stops here. As he admits, he has n o explicit account of what makes some kinds natural and others not, and so at this point simply rests his case on intuitions.

Other anti-reductionists have adopted a rather more sophisticated approach. Instead of resting their case on intuitions as to whether the disjunction (P1 v . . . Pn) is a natural kind, they have pointed out that the relevant disjunction might be infinitely long, or indeed mightn't even be recursively specifiable (cf. Hellman and Thompson, 1975), and that therefore the que stion of any reductive explanation of high-level laws in terms of lower-level ones doesn't even arise.

This more sophisticated line is certainly of technical interest. But I would like to point out that the argument of this chapter adds weight t o Fodor's view that even a finite disjunction of physical states can fail to qualify as a reduction (even though it disagrees with Fodor on the extent to which such reductions are needed). For the argument of this chapter suggests that such a finite disjunction of different physical states ought not to count as a reduction, whenever the picture it leaves us with at the reducing level is physically incredible.

Suppose, to return to my earlier illustration, that the appearance of inflamed ankles a nd knees as a result of eating reheated brussels sprouts has a name -- say, brussitis. And suppose, as before, that different cases of brussitis are due to different physical processes. My point is that the list of such physical processes does n't need to be recursively unspecifiable for us to feel that there is something unsatisfactory about the equation of brussitis with the disjunction of those physical processes. For, as soon as the number of the disjuncts gets above two or three, wew ill judge, quite rightly, that it is incredible that there should be no further explanation of why reheated brussels sprouts always lead to inflamed ankles and knees.

In effect I am suggesting that the notion of a reduction is precisely the notion of an account which shows that nothing incredible is happening at the physical level. Fodor says that a finite disjunction is not a reduction because the physical categorization involved isn't "natural". I am adding the thought that this lack of "naturalness" resides in the fact that such disjunctions are too heterogeneous for it to be plausible that there should be no further explanation for the disjuncts all producing the same effect (a thought which is not available to Fodor himself, given tha t he sees no need for such explanations).

2.7 Teleology and Irreducibility

Despite the argument I have been developing in this chapter, I don't in fact think that psychological categories are reducible to p hysical ones. I think there is a different, non-reductive explanation for why variably realized psychological states often produce uniform physical effects. It is high time I explained how this might work.

By way of an analogy, consider th is example. All domestic water heaters contain thermostatic devices which stop the heating when the water gets hot enough. If we denote the threshhold temperature by R, the thermostat operating by S, and the heating stopping by T, then we have the generalization, applicable to all domestic water heaters, that R -> S -> T. However, there clearly isn't any physical reduction of S here: there are many different kinds of thermostats, with quite different designs and

constitutions, and with nothing physically in common apart from their all turning the heater off when the water gets hot enough.

Even so, there is scarcely much of a puzzle here as to why all the physically different realizations of S produce the common result T. The obvious answer is that water heaters are designed by people not to burn out, and that's why they all contain thermostats that switch off the heating when the water gets hot enough. We can say the mechanisms in water heaters have been selected by the designers in order to switch the water heaters off. That's what the thermosats are there for.

Another example. All vertebrates who breed within a fixed location will act towards invaders of that territory in such a way as to frightena way those invaders. Here let R be the invasion of the territory, S the characteristic behaviour, and T the departure of invaders. Then, plausibly, for such animals, R -> S -> T. Yet there is no physical reduction of S: there is not hing physically in common between all the different forms of territorial behaviour displayed by vertebrates, apart from the fact that they all make intruders go away.

But, once more, there is scarcely anything puzzling here. The obvious explanat ion for the fact that these physically different kinds of behaviour all have the uniform effect of frightening away intruders is that natural selection has favoured those behaviours precisely because they frighten away intruders. As in the previous example, the different physical causes have all been selected in order to produce that effect.

I favour the "aetiological" theory of teleological notions like function, purpose and design. (See Wright, 1973; Millikan, 1989b; Neander, 1991a, 1991 b.) According to this theory, it is appropriate to say that item X has the function of doing Y just in case item X is now present as a result of causing Y.[5] The paradigm for the aetiological theory is the kind of case where X has been nat urally selected by a mechanism which picks out things that cause Y, as in the case of biological evolution by genetic selection. But the aetiological theory can also be extended to cover artefacts like thermostats, and indeed human actions in genera l, since human decision-making can itself be thought of as a mechanism that selects artefacts and actions because they produce certain effects.

Not everybody agrees about the aetiological account of teleology. Some people will want to put scare quotes around words like "purpose" and "design" when they are used in connection with blind mechanisms like genetic natural selection. But we need not spend time on this issue here. For the terminology of "purposes" is not essential to my cent ral point -- namely, that there is nothing puzzling about physically quite different things all having the same effect, if those physical things are all products of some mechanism which selects items because they have that effect. Adherents of the a etiological theory will be able to express this point by saying that there is nothing puzzling about the non-reducibility of some special science's phenomena, if those phenomena are there for a purpose. But others can put scare quotes round "purpose " here if they like.

Let us now consider the specific question of the reducibility of the generalizations of psychology. Here another kind of selection mechanism, different from both biological genetic selection and from intelligent decision-mak ing, becomes relevant. This is selection by individual psychodevelopmental learning. There are

good general reasons for supposing that individual learning, at least in its early stages, must involve some innate tendency to enhance those neural pathways which lead to certain kinds of results, and to discourage neural pathways which lead to other results.[6] In this sense learning is itself a mechanism that selects items because they produce certain results.

As with all selection mechanisms, the items between which this mechanism chooses will be relatively random, depending on such things as idiosyncracies of individual circumstance, linguistic training, knocks on the head, or even on genuinely chance occurrences in the brain. (Compare the "mutations" which are inputs to genetic selection.) From the point of view of learning, the precise physical nature of the relevant items doesn't matter, provided they produce the right kind of effect.

And this, finally , is why there is no reason to expect there to be anything physically in common between two people when they both believe, say, that there is an ice-cream in front of them, even though this state has similar effects on their behaviour. The physical realizations are likely to be different simply because the inputs to each individual's learning mechanism (the "mutations") will be relatively random. But we needn't be puzzled as to how there can be similarity of effects without the physical common ality, for the one thing that the learning mechanism will have ensured that the different states which arise when different people look at an ice-cream will at least share the feature that they will produce appropriate effects in appropriate circumstances (such as reaching out for it when you are hungry).[7]

I earlier mentioned David Lewis's view that mental states are variably realized across species, but uniformly realized within species. The argument of this section corroborates Lewi s's commitment to variable realization across species. But it also suggests that he is wrong to stop there, and that we should expect to find variable realization within species too. Across species we find variable realization of innate mental states because genetic natural selection preserves any genetic mutation with beneficial effects, and such mutations are likely to be different in different species. Entirely analogously, if each individual contains a learning mechanism which preser ves any "physiological mutation" with certain beneficial effects, and if these physiological mutations are different in different individuals, then the upshot will be that even among conspecifics we will find variable physical realizations of acquired men tal states.

Actually, in the case of the mental state that Lewis himself concentrates on, namely, pain, there is a reason to expect uniform realizations within species. For pain is best thought of as part of our learning mechanism, rather than a s the kind of mental ability that this mechanism produces, given that learning selects precisely those mental items that don't cause pain. Since our basic ability to learn isn't itself learnt, but a consequence of our genetic endowment, this is then a reason for thinking that pain, and similar basic mental states like hunger, temperature perception, and so on, are uniformly realized within species.[8]

So far in this section I have indicated a way of understanding how various biological and psychological non-dispositional categories might have uniform effects even though variably realized: such variable causes can have uniform effects in virtue of mechanisms which select items because they have that effect. The corollary, how ever, is that we shouldn't expect non-reducibility in those special sciences where no

such selection mechanisms are to hand.[9] This seems to me to include nearly all special sciences apart from biology and psychology. Perhaps there are som e rudimentary selection mechanisms in some of the social sciences, in the form of economic or social competitition. But there are certainly none in such special physical sciences as meteorology, or chemistry. In any case, the general moral sho uld be clear: special categories that aren't products of selection will be reducible.

Of course, reducibility to physics won't be at issue for sciences whose entities are partly psychologically constituted, like sociology or economics, or partly biologically constituted, like demography or epidemiology. For, even if such sciences are reducible to psychology or biology, the selection-based irreducibility of the latter sciences to physics will block the overall reduction. However, the point remains that the absence of selection mechanisms within such sciences as sociology or demography will imply that such sciences will at least be reducible to their psychological or biological constituents.

This last point is illustrated by Fodor' s (1974) discussion of Gresham's law, the economic principle that bad money drives out good. Fodor stresses the extreme implausibility of any uniform physical realization of such categories as good and bad money. And of course he is right abou t this. But at the same time it is worth noting that there are obvious psychological reductions of these categories, and correspondingly an obvious psychological explanation of Gresham's law. (Money consists of items whcih people exchange for goods and services because they expect others to do the same; such money is good or bad to the extent people believe it will continue to be so exchangable; which is why people will circulate their bad money, and hold onto their good.) Of course these psychological facts won't in turn reduce to physical facts, given the teleological underpinnings of psychology. But the reduction of the economic facts to psychological facts is just what my overall theory predicts, given that there ar e no economic selection mechanisms to provide an alternative explanation of Gresham's law.

2.8 Selectional Explanations

An obvious question raised by the argument of this chapter is the status of selection al explanations themselves. I have argued that in disciplines with a teleological underpinning, such as psychology, we can explain why the disparate physical realizations Pi, of some given special category S, all have the same effect T, by invoking a selection mechanism which picks out Pis precisely because they cause T. Implicit here is the general claim that, in systems of the relevant kind, if any Pi causes T, then that Pi will tend to be preserved. But what now about this general cla im? Is it itself reducible to physics? And, if not, doesn't it raise just the same puzzle about variably realized generalizations having common effects with which I started the paper?

I think that some such selectionist generalizations (if Pi causes T, then Pi is preserved) are physically explainable. Others, however, will be variably realized. But then there will be some more general selection mechanism which in turn explains the existence of the specific selection mechanisms which pick things that cause T. What about the generalization implicit in this last explanation (if a selection mechanism picks Pis that cause T, then this selection mechanism will be preserved)? Well, either this

generalization will be physic ally explainable, or it will be due to a yet more general selection mechanism which . . . is physically explainable.

So I would say that we can explain patterns that aren't themselves physically explainable in terms of selection mechanisms which are, or at least in terms of selection mechanisms whose selection is physically explainable, or . . . and so on. We can have hierarchies of selection mechanisms, with variable realizations at each level until the last. But the last level should alw ays offer a uniform physical story, for until we have such a uniform physical story the explanatory buck with which I been concerned in this paper won't have stopped.

Let me give an illustration of the simplest case, where some variably realized patte rn S -> T is explained by a physically uniform selection mechanism. Consider some simple biological organism which is capable of learning how to get rid of some given painful stimulus. Different individuals in this species will learn physicall y different ways getting rid of the pain. Then, if we think of the avoidance behaviour as S, its different realizations as the Pis, and the disappearance of pain as T, we will have the generalization S -> T, and this will be variably realized by the different Pis. This is the kind of variable generalization that this paper has argued to be prima facie puzzling. In the example at hand we can remove this puzzlement by invoking the learning process which ensures that, if any Pi causes T, that Pi will be preserved. The current issue, however, is what kind of explanation we can give of this general selectionist fact. But now recall a point made in the last section, that pain and associated learning mechanisms are likely to be innate and so uniformly realized within any given species. If this is right, then a uniform physical explanation for the selectionist generalization will be available. The story will no doubt be complicated. Nevertheless, to postulate that the lea rning mechanism is uniformly realized throughout the species is precisely to postulate some physically uniform feedback mechanism which is triggered by the disappearance of pain, and which then operates to preserve whatever physical behaviour caused thatd isappearance.

Does this mean that in this kind of case the original behavioural generalization S -> T will be reducible to physics after all? Only in an extended sense. There is still no uniform physical explanation of why the behaviour S generally gets rid of pain, for S is still variably realized at the physical level. Rather, what we can explain physically is why each individual, on receipt of the painful stimulus, performs some bit of behaviour which gets rid of the pain. In effect, what the selectionist story allows us to explain is not so much why the behaviour has the effect it does, but rather why each individual is disposed to some bit of behaviour with that effect. This explanation does, it is true, imply that all the behaviours in question have the effect they do; but it doesn't do this by identifying a uniform physical reduction for those behaviours; rather it switches to a broader context and instead gives a uniform physical account f or each individual having some behaviour with that effect to start with.

Now for a more complicated case. Suppose again that some group of animals have learned some common but variably realized behaviour, but not now because of their innate tend ency to avoid pain, but rather because they have all acquired a common desire in virtue of their their common experience. Suppose, for example, they have all learned to like bananas. And suppose that, as a result of having this desire, they ha ve all learned ways (though not necessarily the same ways) of doing such things as

getting bananas down from trees. Now in this case there won't be any uniform physical explanation for their all acquiring such behaviour. For, if the desires fo r bananas are themselves acquired by learning, it is unlikely that those desires will themselves have a uniform physical realization, and so unlikely that the feedback mechanism responsible for learning how to get bananas down from trees will be physicall y uniform across the different individuals.

Here we need to shift to a yet wider context, and focus on how the desires for bananas were acquired in thie first place. At which point we will presumably want to tell a story about a mechanism which selects states (namely, desires) which will cause, and help develop, behaviour which will yield results, such as bananas, which in the organism's experience have been associated with pleasure or the avoidance of pain -- where this selection mechanism is u niformly realized across the species. And this will then, as before, offer a uniform physical explanation, not of how the movements each animal performs lead to bananas (for these movements are physically non-uniform), nor even of how each animal ha s learned some bit of behaviour which gets bananas (these learning processes are physically non-uniform too), but rather of why each animal has acquired a state which disposes it to learn some bit of behaviour which will get it bananas.

We could go on . Generalizations about suitably experienced individuals seeking out bananas will hold good across species, as well as within them. But this then opens up the possibility that the associated innate mechanism for acquiring desires will be diffe rently physically realized in different individuals who alike seek out bananas, thus underming the physical uniformity of the story told in the last paragraph. But then we can widen the context even further, and appeal to intergenerational genetic s election, which will then explain these variably realized innate learning mechanisms as themselves selected by the physically uniform process which preserves things which cause survival and the replication of genetic DNA.

I am not of course suggesting that special scientists need to go into all this every time they appeal to some variably realized special generalization in explaining something. The idea that you must explain everything you use in an explanation is obviously self-defeating. Nevertheless, on the metaphysical level, as opposed to the methodological level, it is worth knowing that if we widen the context enough we can in principle always show that any variably realized special generalization is the upshot of some uniform physi cal process. For if such physical explanations weren't in principle available, it would be incredible that such variably realized generalizations should be true.

1. Functionalism per se can be defined in a way that does not require second-order causal roles to be realized by physical states. However, I have already argued for physicalism, in the last chapter. So I shall henceforth understand "functionalism" to stand for the narrower doctrine which does specify physical realizations.

2. It will ease the exposition if I can assume for the moment that functionalism includes the thesis of variable realizability. So for now functionalism is not just a thesis about the meanings of special terms, but that plus the denial of any typ e reducibility of the special to the physical. I shall return to this point in section 2.4 below.

3. Searle (1985) agrees with this moral, but takes it to provide a reductio of the possibility of special scientific laws: that is, he argues that, if there were special laws, then the categories of the special sciences would have to be reducible to physics; and so, since the categories of the special sciences clearly aren't reducible to physics, there can't be any special laws.

4. I am here thinking primarily of reducibility to intrinsic physical characteristics of red objects, such as the molecular constitution of their surfaces. (Cf. Smith, 1987). But the point also applies (pace Smith) to the reducibility of redness t o such relational physical characteristics as transmitting certain wavelengths of light in certain sorts of illumination; maybe there isn't even anything in common about the relational physical properties of red things, apart from the fact they make normal people see red. Indeed, there arguably isn't even any immediate reason for supposing there must be something physically in common even between the way different people respond perceptually to red objects: couldn't each person learn som e physical way of reliably responding to what everybody else called "red", but with different people doing this in physically different ways? But perhaps this last extreme anti-reductionsist conjecture is in tension with the ability of different peo ple to agree in identifying previously unobserved objects as red, even though redness is a secondary quality whose instances have nothing in common except their ability to make people experience red.

5. It is unfortunate that "function" and its cognat es are used both for the teleological notion of causes that are explained by their effects, and for the definitional notion of concepts that can be defined as elements in a structure of causes and effects. The two ideas are quite distinct.

6. Se e Chapter 3 of Daniel Dennett's Content and Consciousness (1969) for general reasons why learning must work like this, and recent "connectionist" models of pattern recognition for specific illustrations. In particular, see Andy Clark (1989, pp 12-13) for an instance of how similar teaching can train neural nets with random initial conditions into the same (high-level) structures.

7. I am in effect arguing that the categories of non-reducible special sciences must have purposive functions, to exp lain their non-reduciblity. Lycan (19xx) and Sober (19yy) have argued similarly that functionalist theories of mind need to be supplemented by teleology, lest too many systems count as minds: their idea is that not anything with a certain caus al structure should count as a mind, just those systems which are designed to have that structure. However, while the conclusions are similar, it is worth noting that the argument I have given has both wider scope and greater force. It argues that t eleology is needed not just in functionalist accounts of mind, but in any law-governed non-reducible realm. And the rationale is not just that the functionalist theory of mental concepts would founder without teleology, but that there wouldn't be an y non-reducible special laws without teleology.

8. This point might seem open to dispute. If, within a species, two physically different genes produce exactly the same phenotypic effects in different ways, then both genes will be equally favoure d by natural selection. Indeed, short of knowledge of their detailed DNA structure, two such genotypes are likely be counted as one by biologists. This suggests the possibility of two physically different types of pain within one species. ; However, it is empirically unlikely that variant physical

components in such a complex mechanism as pain would ever have exactly the same selection-relevant effects. This case is different from the case of different species, for in different speci es natural selection will favour different physical bases for roughly the same job as long as they have roughly similar effects. But within an interbreeding species there will be direct competition between such physical alternatives, and so such alt ernatives will only both be preserved if they have an exact selective equivalence.

9. Why should appeal to selection mechanisms be the only way of explaining away non-reducibility? But what else could account for the fact that physically dispara te items have the same effect, except some mechanism that picks them out because they have that effect?