EUGEN FISCHER
Ludwig-Maximilians-Universität München

Unfair to Physiology

It is often felt that there is an 'explanatory gap' between the physical and the mental. In part, this is due to the idea that physiology can explain at best an organism's behaviour, outward and inner, but not the conscious experiences that accompany that behaviour. By a novel combination of approaches due to G.E.Moore and Ludwig Wittgenstein, respectively, the paper refutes this idea, and seeks to pin down and dispel a peculiar sense of intellectual unease this refutation may leave us with. To refute the idea, the paper clarifies it by confrontation with an actual example of psychophysical explanation of perceptual experience. This reveals that the idea relies on a prejudice about physiological practice. Then the paper explores some peculiar ways in which this prejudice may survive its refutation. This is to bring out that such explanations of experience as are actually offered by contemporary psychophysics explain nothing less than what they purport to explain; and that these achievements are not, in some peculiar way, more remarkable than equally clever physical explanations of other phenomena.

Keywords: experience, explanation, explanatory gap, perception, philosophical therapy

---

'How is it that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue?' philosophers wondered ever since T.H.Huxley brought up the question. Finding themselves at a loss for an answer, some philosophers concluded that 'we have no conception of our physical ... nature that allows us to understand how it could explain our subjective experience' (Block, 1994, 210). It is, perhaps, not quite clear what *kind* of 'conception' and 'understanding' they have in mind when drawing this conclusion. But it is probably fair to say that most of them feel the need for such a 'conception' for two quite different reasons, viz., because they feel puzzled by two distinct questions: 'How can facts about our physical nature explain why we have any such thing as experience at all?' And 'How can such facts explain why we have the experiences we have, in the circumstances in which we have them?' The present paper is

devoted to the latter question, rendered slightly more tractable as: 'How can physiology, the study of the physical and chemical on-goings in our bodies, explain why we have the experiences we have, in the circumstances in which we have them?'

Philosophers' How can...?-questions are often motivated by appealing objections to the effect that ... can*not*. Some such questions are best 'answered' or, rather, coped with simply by refuting the pertinent objections. The present paper aims to examine and refute what I take to be the most natural and persuasive objection to the effect that apparent physiological explanations of experience cannot explain what they purport to explain. The refutation of this objection is intended to remove our main motivation for wondering 'How can physiology explain why we have the experiences we have, in the circumstances in which we have them?' If successful, this refutation will help us to attain one pertinent sort of 'understanding' of 'how our physical nature could explain our subjective experience'. To achieve it, the paper employs a novel strategy combining two familiar tactics, due to G.E.Moore and Ludwig Wittgenstein, respectively. This exercise has a two-fold aim: It is to reduce the intellectual vertigo induced by Huxley's unclear but perplexing question; and it is to remind us of the virtues of two philosophical approaches that are fading from practitioners' memory in spite of their manifest potential.[1]

1. The Natural Objection

This is a natural idea: To the physiologist an organism, rat, bat, or human, quite simply is a 'meat-machine' (to borrow a vulgar but suggestive term from Marvin Minsky); to him, the organism is but an intricate complex of muscles and sinews, tissue and blood-vessels, of, ultimately, cells in which various physical and chemical processes take place. The physiologist explains the behaviour of the organism by reference to the behaviour of its organs, and this, in turn, in terms of the physical and chemical processes they are involved in or driven by, by reference to the laws of chemistry and physics. But then he will, at best, explain why the wound bleeds and why the body writhes, or why the subject can discriminate two stimuli. Why the subject *feels* pain or why the stimuli *look* different to him is bound to remain a mystery. Physiology can explain why organisms and their parts behave the way they do. It can explain an organism's behaviour, outward and inner. But not the

conscious experiences that go with that behaviour. Let this be our '*intuitive objection*'.

Functionalism makes this complaint appear naive. Of course, organisms are but 'meat-machines' to the physiologist. But, put abstractly, he thus considers the organisms as made up of various bits and pieces in various physical and chemical states that confer certain causal properties on them. So there is at least no conceptual obstacle to physiological explanations of mental states amenable to functionalist analysis: If an internal state individuated in the physiologist's terms happens to stand in just that pattern of causal relations to environmental input, other internal states, and behavioural output, definitive of a certain (say) experiential state, then that physiological state will be a realisation of this experiential state, and the physiological explanation of the former will also explain why the organism is in the latter, i.e., why it has that experience. But if one found the intuitive objection appealing, one will remain puzzled even if one grants a functionalist analysis and assumes that the facts of physiology mesh with it: On the present explication, what the physiological explanation of an organism's experience actually explains is simply why it is in the internal state that stands in such-and-such causal relations to environmental input, behavioural output, and other internal states. But why does it also feel for the organism the way it does to be in that internal state? Physiology may be able to explain why, under certain circumstances, we are in internal states with the functional role of, say, headaches or auditory states. But it is doomed to leave unexplained why we feel a dull, thumping pain or hear a certain buzzing sound when we are in those internal states. Let this be the '*sophisticated objection*'.

If you found the intuitive objection intuitive, functionalist theories and their predecessors will all leave you with some such unease as translates into a version of the present 'sophisticated' objection. I.e.: If prior to philosophical theorising we are tempted to object that physiology can merely explain behaviour, the sophistication introduced by philosophical theory won't rid us of the feeling that any purported physiological explanation of experience is bound to 'leave out' the very thing it purports to explain. Hence we seem well advised not to rush into that sophistication, but to attend first to the simple-minded but powerful idea that induces that feeling already prior to any explicit philosophical theorising, viz., to our intuitive objection. The aim of this paper thus is to refute *this* objection.

The idea it vents is, however, not the only source of the feeling that there is an 'explanatory gap'. This feeling is, rather, due to a variety of quite distinct, and

mutually reinforcing, confusions and prejudices. By refuting our intuitive objection, we will free ourselves of one of the latter, no more and no less. This will not rid us of any such thing as 'the problem of the explanatory gap', but will do all that can sensibly be done within the scope of a single paper: It will make a little shorter the garden path led up which we have come to feel that there is some such problem. And it will thereby show us what kind of thing can help to shorten the path further.

The present aim confronts us with an elusive opponent: Many of those attracted by the 'intuitive objection' find it so simple-minded that they are embarrassed to put it in print, venting it only in oral discussion, e.g., when pressed to put in plain English the motives that have them take part and interest in apparently more sophisticated and printable argument. (For written evidence one therefore best turns to conference minutes such as, e.g., Bock, 1993.) As a result, the objection is generally denied serious attention and escapes the thorough criticism in the absence of which it arguably continues to impress us. This gives the objection an influence both less visible and more tenacious than that of the various more prominently published versions of its sophisticated cousin. We should, therefore, bring it out into the open and face squarely up to it. This paper is to do just that, and to cope with the objection unearthed, in two steps: To refute it, the paper clarifies the compelling but unclear objection to a point at which it becomes amenable to sober confrontation with hard facts, and confronts it with facts uncongenial to it. Then the paper pins down, and tries to dispel, a peculiar sort of intellectual unease with which this exercise may leave us.

2. An Actual Example

Our first task thus is to transform the philosophical poetry of our intuitive objection into tractable prose. To do so, let's confront the intuition this objection articulates with a concrete example of the sort of explanation of perception contemporary physiology actually offers, and see just what objection against it our intuition has us make. So consider these elements of psychophysical explanation of colour-vision.[1] (To get the hang of this explanation, we will have to go into a little more detail than is explicitly seized on by the subsequent argument.)

[1] For a fuller account of these see: Clark 1993, ch. 2. It will be readily apparent where I diverge from Clark's account.

Light quanta produce their effects on receptor cells by isomerising (i.e., changing the configuration of) molecules of the photopigment in those cells: This affects the ion permeability of the cell membrane and thereby alters the electrical potential across the membrane. Such cell responses depend exclusively upon the number of quanta absorbed by the receptor. This number depends on two factors: the number of incoming quanta, and the proportion of quanta absorbed by the pigment. For all pigments, the latter differs with wavelength. This means that receptor response is, ultimately, a function of intensity and wavelength of the incoming light. Hence differences in the former can be made up by differences in the latter, and *vice versa*. So individual receptors respond the same to indefinitely many stimuli of different wavelength and different intensity.

The human retina contains different kinds of receptor cells. In particular, it contains three distinct types of cones with pigment with different absorption spectra, peaking at short, medium, and long wavelengths, respectively. One and the same stimulus will therefore have different effects on these receptors. Between cells excited by these different kinds of cones there are inhibitory connections, resulting in subtraction of the output of some such cells from that of others. The responses of the receptor cells exciting them is roughly proportionate to the logarithm in stimulus intensity. Subtraction of logarithms is equivalent to division. The physiological make-up of the visual system thus allows us to regard it as forming ratios among the outputs of the three distinct receptor systems. Stimuli that cause the same effects on individual receptors may have distinguishable effects at this later stage of visual processing. Conversely, stimuli that had different effects on individual receptors may at this stage have the same effect. Indeed, light of the same wavelength but different intensity will cause different effects on the receptors, namely, absorption of different numbers of quanta and, hence, different firing rates. But the ratio of the numbers of quanta absorbed by cones of the three different systems will be the same, over a wide range of intensities. At this stage, and only at this stage, neural processing becomes sensitive to wavelength.

The different responses possible for the system in these early stages can be represented within a three-dimensional co-ordinate system or *wavelength mixture-space*. Each axis indicates the number of quanta absorbed by one of the three types of receptor. Each point in the space thus defined represents a distinct initial response. The computation of ratios then falls into place as the formation of vectors: The responses to light of the same wavelength but of varying intensity

will be represented by points along a vector in that space. Its length will represent the intensity, its direction the wavelength composition of the stimulus. To construct an individual's wavelength mixture space, we just need to know the absorption-spectra of the pigments in his three types of receptors.

The individual's colour-vision can then be investigated by means of a '*colorimeter*': a device that can present two distinct mixtures of wavelengths to an observer. Each mix is presented as half the field visible through an eye-piece. In one half of the visual field we have the test-stimulus whose wavelength mixture is being controlled by the experimenter. In the other half the mix consists of a few 'reference' wavelengths whose relative intensities can be adjusted by the observer. The observer is asked to adjust the relative intensities of the reference wavelengths until both sides of his visual field, i.e., both wavelength mixtures, look the same to him. His sincere and careful assertion, at a certain setting, that both sides (i.e., wavelength mixtures) now look the same to him, is the last word on whether the two stimuli presented *match*.

Testing for matches requires care: Upon prolonged exposure to given stimuli, or upon comparison to others, subjects typically learn to distinguish stimuli that previously looked the same to them. Time and further comparisons will therefore yield more precise matches. To complicate matters further, we sometimes feel unable to tell whether two things look the same, when differences get very small: One moment we are inclined to say one thing, only to prefer the other, in the next. To get a determinate response even in such cases, experimenters can revert to a forced-choice task: They switch the test- and another stimulus around in a random fashion, and ask the subject to identify the previously designated test-stimulus. The subject is to respond (to press the appropriate button) even if he feels unable to tell: If unsure, he just is to guess.

> The two stimuli are *indiscriminable* for the subject iff he is unable to identify the test-stimulus at a significantly higher than random rate, over a series of trials.

(What is to count as a 'significantly' higher rate is, to a certain extent, a matter of stipulation, viz., in statistical terms, of the desired level of confidence.)

We then find that both matching and indiscriminability stand in a meaningful relation to the effects stimuli have on early stages of subjects' visual processing: All stimuli that have the same effects on the subject's visual system, i.e., effects that are represented by the same region in wavelength mixture space are judged to look the same, under specific laboratory conditions. And for each subject there is a reasonable level of confidence at which two stimuli are indiscriminable under

those conditions if they have effects on the subject's visual system represented by the same point in wavelength mixture space. The construction of such a space therefore allows us to predict which stimuli a subject will find to match or be indiscriminable, under laboratory conditions: Each point in the three-dimensional space can be reached by addition of three vectors, in indefinitely many ways. The physiologist predicts a match whenever vector addition yields the same result. And, under laboratory conditions, these predictions are always borne out by the matching judgements of experimental subjects. The physiologist will then explain the match between two stimuli of different spectral composition and intensity by reference to the identity of the effects they cause at this early intermediate stage of neural processing. He will say that the two different stimuli are indistinguishable for the subject because they have such identical effects on his visual system. Indeed, he will say that because they have the same effects at this particular stage of the visual processing, the two stimuli look the same to the subject.

3. The Objection Refined

We are then tempted to sum up the physiologists' procedure like this:[2] To explain why two stimuli x and y, under certain experimentally replicable conditions, look the same to a subject, they take two steps:

(1) They show that, at some stage of the visual processing, certain cells C are crucially involved, and that both x and y cause cells C to have some property Q. Then

(2) They show that stimuli that cause C to have Q are indiscriminable (from the very same stimuli).

On this basis, they predict that, under the pertinent experimental conditions, any two stimuli will look the same to a subject if they both cause C to have Q. And explain that, under those conditions, x and y look the same to the subject because both cause C to have Q.

This sort of explanation is unabashedly *non-reductive*: It doesn't even try to give us the kind of 'strong theoretical understanding' of why two stimuli that both cause cells C to have property Q 'must' look the same to the subject that some philosophers of mind hanker after. So it does *not* confront us with a version of the

[2] Clark, e.g., succumbs to this temptation. Note he also adds on step (3) below. See his 1993, 27-8.

'problem of the explanatory gap' that has recently been raised about the possibility of *reductive* explanations of experience that provide such strong theoretical understanding (see, e.g, Chalmers, 1996, chs. 2-3). But our initial intuitive objection gets us into a *related* problem, as it has us wonder: Does the present explanation give us even a suitably 'weak theoretical understanding' of the subject's experience — or only of his behavioural dispositions?

A brief glance at (1) and (2) suggests the latter: What these steps appear, at best, to explain is why the stimuli in question are *indiscriminable* to the subject. The conclusion that they therefore *look the same* to the subject, we feel, requires a further step:

(3) Physiologists eventually need to draw the inference:
  (I1) If two stimuli are, for a subject, indiscriminable (from the very same stimuli), they look the same to that subject.

This step we then feel uncomfortable about, as it seems to miraculously get us from the realm of mere behaviour to that of full-blown experience. And, indeed, it appears the crucial inference (I1) cannot be justified by the physiologist: We started out from the idea that he considers his subjects as 'meat-machines', as organisms whose workings he explains by detailing the workings of their several parts in physical and chemical terms and explaining them by invoking the laws of physics and chemistry. But then all he is, at best, able to predict is the twitching of the fingers working the buttons in the course of testing for the capacity to re-identify a 'target-stimulus' over a sequence of trials. Another look at the physiological explanation under scrutiny then seems to reveal that this is precisely the sort of thing the physiologist predicts in predicting which stimuli will be indiscriminable for the subject: 'Discriminability is a purely physical notion... Its import is just that a particular statistical relation obtains among stimulus events and choices' (Clark, 1993, 50). This notion appears to be applicable not only to human subjects but every bit as much to robots and suitable detection-devices — who fail to enjoy any experience whatsoever. Inference (I1) is clearly not valid: Inanimate counter-examples are obvious.

A brief story brings out that the inference cannot be saved by restricting the domain to animate beings, or even to humans:[3] A mishap befalls Dan, who volunteered to serve as a subject in a forced-choice experiment on colour vision of

[3] To my knowledge, this story, recently re-told by John Searle, is originally due to Kirk (1974).

the form described above. As he tells us moments later with all signs of alarm: 'I lost my sight while peering down the eye-piece of the colorimeter. To make matters worse, I simultaneously lost control over my own movements; while I would have very much liked to jump up and ask for immediate medical assistance, I just could not get my body to do that. Indeed, I felt that my eyes remained glued to the eye-piece, and that my fingers twitched, thus pushing the buttons at regular intervals — an experience a lot more distressing than feeling the knee jerk in a reflex test.' If we took this extraordinary story to be the sincere and careful report of a linguistically competent subject, we would not want to say that Dan ever saw whether the test-stimulus was left or right, but that he did not see a thing. But of course his fingers did push the pertinent buttons as the stimuli were set to various intensities, and indiscriminability, understood as a 'purely physical notion', will apply, if his button-pushing responses track the test-stimulus with a significantly higher than random rate of success. We can thus conceive also of human counter-examples to (I1). That inference is not logically valid. Nor can the physiologists justify their step by induction: All they consider are the subject's discrimination-behaviour and the concurrent internal physiological goings-on. Study of the organism's internal behaviour will yield not (I1) but such insight into its causal workings as gained in step (1). Observation of its outward behaviour yields the verdict of whether the stimuli are indiscriminable to the subject — an instantiation of the antecedent of (I1), not the complete conditional. And observation of their co-occurrence, finally, will yield not (I1) but rather the sort of conditional established by step (2): If stimuli x and y both cause cells C to be Q, they are indiscriminable.

Once neither deduction nor induction have proven up to the job at hand, inference to the best explanation is the obvious candidate: The ascription of perceptual experience provides the best explanation for the subject's discrimination-behaviour; he fails to identify the test-stimulus at a higher than random rate because the test and the 'foil' stimulus look the same to him. But whatever else this is, it is not a *physiological* explanation like the explanation that x and y are indiscriminable for the subject (or that x and y look the same to him) because both stimuli cause cells C to have property Q. The physiologists could not invoke (I1) as an inference to the best explanation without making use of 'explanations' that are not part of their subject, thus rendering dubious their claim to have provided a *physiological* explanation.

This leaves us with this objection: 'Step (1) establishes a truth about the behaviour of certain cells. Step (2) reveals how this hangs together with the

behavioural dispositions of the subject. To get from these findings about the organism's internal and outward behaviour to conclusions about what things look like to the subject, physiologists need some such inference as drawn in step (3). And physiologists cannot justify this inference, (I1), at any rate not unless they invoke explanations that are not part of their subject. If they respect the limitations of their discipline, all they actually manage to predict and explain (in whatever weak sense their unabashedly non-reductive explanations do "explain" anything) are the subject's behavioural dispositions.' Let this be our '*specific objection*'. It articulates, in more or less tractable prose, what our intuitive complaint, that physiology 'can only explain behaviour', comes down to in a particular case, when confronted with an actual example of such explanation.

4. The Objection Refuted

This specific objection has two parts: We construed the present explanation as relying on a certain inference, viz. (I1), thus challenging physiologists to justify it. And we argued that this challenge cannot be met, at any rate not within the confines of physiology. Familiar philosophical theories of mind would have us look for different strategies to meet this challenge. A second look at our present explanation suggests that physiologists do not face it, in the first place: Our reconstruction of their proceeding is simply inaccurate. With remarkable ease, we swept part of their results under the carpet: They test for both indiscriminability *and* matching, and found *both* to stand in a meaningful relation to the effects stimuli have on a certain early stage of neural processing. But we immediately focused on their finding concerning indiscriminability (apparently a cosily physical notion), and simply ignored their use of matching (regarding it, perhaps, as a bit of a dirty short-cut).

Experimenters, however, usually test for matching, not indiscriminability (a very time consuming procedure), and require tests for the latter only in very specific contexts: When seeking to infer the number of receptor systems operative under certain conditions from colorimeter-data, e.g., physiologists found it necessary to start out from discrimination-data, rather than reports of matching. Here, they sought to infer identity of stimuli's effects on the subject's retinal receptor cells from his discrimination-record, and argued since Cornsweet (1970, 217-19) that the stimuli must have different effects if the subject's behaviour is in *any* way sensitive to the difference between them. But when testing predictions of which

stimuli will look the same to a subject, they test for matching: They then record the intensities to which the subject sets the light when asked to adjust them until it looks the same as the test-stimulus. And they treat the subject's sincere and careful verdict as the final word on whether the two stimuli presented match.

I.e.: Physiologists assume their subjects are linguistically competent, co-operative, and careful. They then elicit responses from them which they treat as sincere reports of whether the stimuli presented look the same to the subject. When they judge the subject sufficiently careful, they infer from his report a conclusion as to whether the stimuli looked the same to him or not. They thus draw the inference:

(I2) If a linguistically competent subject reports sincerely and carefully 'x and y look the same to me now', then x and y look the same to him, at that time.

Physiologists then find that subjects report a match just in case the given stimuli cause cells C to have property Q. On the basis of this finding, they make a prediction. The prediction is obviously not that subjects will report a match whenever two stimuli cause their cells C to have property Q: In almost all cases in which stimuli have such effects on our retinal processing, we don't say anything about how things look to us. Physiologists might of course predict that when two stimuli cause a subject's cells C to have property Q, the subject will report a match, if co-operative, sincere, careful, linguistically competent, and asked how things look to him at that time. But they don't. Instead, they predict that when two stimuli cause a subject's cells C to have property Q, they look the same to him. In virtue of (I2), these predictions are taken to be confirmed (or disconfirmed) by reports of matches subjects make when their cells C are being caused to have property Q. And if they have been confirmed sufficiently well, physiologists endorse the explanation that the stimuli in question look the same to a subject because they all cause cells C to have property Q.

This means that two of the assumptions on which we based our initial objection are simply false: *Pace* (2), physiologists test for matching, rather than indiscriminability, and show not that stimuli that cause C to have Q are indiscriminable for the subject but that such stimuli are found to match. So, *pace* (3), they are in no need of any inference from discrimination-data to conclusions about how stimuli look to subjects: They can rely instead on inference (I2). But then they are under no obligation to justify the rather different inference (I1) we attributed to them. They don't face the challenge we believed to confront them.

What is more, the inference they do employ, instead, is one we were happy to countenance: While sceptical about (I1), we did countenance inferences from subjects' reports — in conjunction with the assumptions of sincerity, care, and linguistic competence — to conclusions about how things look to them: In framing our specific objection, we drew such inferences ourselves, viz., from the story told by Dan. But then we have to accept (I2). Which obliges us to grant that the physiologists can predict when two stimuli look the same to a subject, and do provide one sort of explanation of this fact about his perception — rather than his behavioural dispositions. I.e.: We have to drop our specific objection.

This specific objection is all we have explicitly argued against so far. All we have shown so far is that the one particular psychophysical explanation we considered does not 'merely explain behaviour'. But as a single counter-example suffices to disprove a general claim, this is enough to refute our sweeping initial objection that 'physiology can merely explain behaviour', by which we presumably meant that *no* purported physiological explanation of a fact about our experience can explain more than a fact about our behaviour.

So let's take stock: To get more clear on our initial objection, that 'physiology can only explain behaviour, but not experience' (voiced in section 1), we considered (in section 2) an actual example of physiological explanation of perceptual experience, and spelled out (in section 3) what our initial complaint comes down to in this particular case, when confronted with the actual example considered. Once we had translated our initial complaint into the more tractable prose of the specific objection we thus obtained, we could refute this objection (in the present section 4) by doing no more than pointing out some salient features of the concrete example of physiological explanation of perception we considered: All we just did was to remind ourselves that (i) physiologists test for matching and (ii) employ inference (I2), and to point out what is then obvious: that the argument leading up to our objection is faulty, as (i) and (ii) disprove our assumptions (2) and (3), respectively; and that (I2) is the sort of inference we used ourselves in framing that objection, so that our objection commits us to this inference, and we have to accept the physiologists' explanation as one of a fact about our perceptual experience, rather than about our behavioural dispositions.

So, in a nut-shell: We spelled out what our intuitive objection comes down to in a particular experimental context, and were then able to refute it by confrontation with obvious facts about the particular scientific practice we considered. This is an approach with a pedigree — G.E.Moore liked to employ a similar tactic: to

spell out what philosophers' theoretical contentions come down to in plain English, so as to be able to refute them by confrontation with obvious facts acknowledged by common sense, as implicit in ordinary language (cp. White, 1958, chs.1-3). The only pertinent difference is that we considered scientific, where Moore attended to 'ordinary' practice.

5. The Persistence of Prejudice

Once we have seen *where*, it is also easy to make out *why* we went wrong: We got things wrong in making assumptions (2) and (3). We made these assumptions because we ignored the tests for matching. So why did we ignore these tests, why did we focus exclusively on the forced-choice experiments testing for discriminability? Presumably, we did so because we took for granted that physiologists make do with 'purely physical notions' — like discriminability, and unlike matching. This assumption, that physiologists 'consider their subjects as "meat-machines"', struck us as so natural an idea that we invoked it when first outlining both our 'intuitive' and our 'sophisticated objection', at the very outset. But, as it happens, the actual experimental practice we then went on to consider simply is not in line with our 'natural idea': Physiologists don't always treat the behaviour of their subjects as that of 'meat-machines', but at least sometimes regard their utterances as sincere reports, i.e., as intentional actions — to which they apply other than purely physical notions.

However, the natural idea thus revealed to be a prejudice may well, in a sense, prevail over the established facts; it may exert a peculiarly tenacious influence, even once we know it to be false. One characteristic reaction is this: We acknowledge that in the course of the actual experimental work considered physiologists do treat their subjects as agents rather than 'meat-machines' — and yet feel somehow uneasy about this: In spite of our acknowledgement, we are, as likely as not, inclined to regard the psychophysical explanation considered as either a sham or a particularly remarkable achievement. More particularly, we either find it a cheap trick (or a 'dirty short-cut') that physiologists test for matching, rather than indiscriminability ('considered as a purely physical notion'), or find it particularly remarkable that they are not restricted to the latter. We are inclined either to object, 'But surely all that physiologists have the right to test for is indiscriminability!' or to marvel, 'But how can physiologists test for matching?'

Both inclinations are odd. We have already provided the answer to the question. — How they can do it? Well, they ask the subject, Mr. X, to peer down a colorimeter, to adjust the settings until both sides look the same, they then explain how to adjust them, etc. And we don't really know what to mean by the objection: In which sense do we want to claim the physiologists 'lack the right' to proceed in this way? Of course they need not subscribe to unwarranted assumptions in doing so: Suppose they choose subjects who know English and are known to be careful and sincere, e.g., their own graduate students, and make sure there is enough time for careful consideration, etc. This clearly warrants the assumption that the utterances in question are careful and sincere reports of matching. So do we mean, instead, that the inference (I2) to the conclusion that both sides look the same to the subject is unwarranted, even if the premises are not? Hardly — as we noted above, we had no qualms about relying on this very inference, ourselves.

So is our complaint rather that while those assumptions and inferences are warranted, *physiologists* are not entitled to them, *qua* physiologists? Unless seized by a grand Hegelian mood, we might feel a bit uneasy about this complaint: After all, physiologists presumably know best what makes for respectable physiology, viz., which terms and assumptions they are entitled to and which not. But, even so, some of us might want to press on: 'That a subject's report is careful and sincere is a *psychological* assumption. The same goes for the assumption (supporting I2) that such reports of matching are true under the pertinent laboratory conditions. Hence the physiologists under consideration aren't providing a *physiological* explanation.'[4] This, however, is an obvious *non-sequitur*: What is established by means of those assumptions and inferences is the fact that both sides look the same to Mr.X, at settings x,y,z. And this may well be labelled a *psychological* fact. But this is of course the fact *to be* explained, whose present explanation invokes but a paradigmatic fact of *physiology*: that the stimuli on both sides cause cells C to have property Q. So we are provided with the very thing we were promised: a physiological explanation of a psychological fact about our perceptual experience.

But still we feel an unease we want to put into the words 'All that physiologists really have a right to test for is indiscriminability' — even though we cannot spell out a complaint we would want to maintain on a moment's reflection, and are, in particular, unable to explain what sort of 'right' we mean. Our prejudice thus persists in rather curious forms: It translates into an unease that has us press an

[4] For this objection I am indebted to an anonymous referee.

objection whose meaning we cannot explain, or into a sense of wonder that has us marvel at a question to which we already know the answer. To stress: Even once we have rejected it as false, our prejudice lingers on as an odd sense of unease or wonder.

This disquieting phenomenon was, to my knowledge, first identified and taken seriously by Ludwig Wittgenstein, in stressing, e.g., that in philosophy there is a peculiar difficulty unlike 'the intellectual difficulty of the sciences', viz., 'the difficulty of a change of attitude. Resistances of the *will* must be overcome' (Wittgenstein, 1993, 406). Let's try to overcome the particular 'resistance of the will' at hand, viz., our resistance to granting physiologists the right to test for matching (rather than indiscriminability), that survives the rational rejection of our prejudice that they 'treat their subjects as meat-machines'.

6. The Prejudice Dislodged

It will help to get more clear on what we mean when we vent this prejudice. Our rough-and-ready idea is that physiologists describe their subjects in terms we could conceive to apply every bit as well to lab-rats — or even to robots. So here is a straightforward explication of what *this* might come down to: Let's distinguish between '*neutral*' and 'action-descriptions'. The former are applicable to outward and inward behaviour regardless of whether it can be characterised as action, in particular, regardless of whether it is under the agent's direct control and of whether it makes sense to speak of it as being performed with some intention. Let's use these two features as 'marks of action' (for a fuller exposition, see Rundle, 1997, ch.6). And let's call descriptions of behaviour in terms that imply that the behaviour in question possesses these two marks '*action-descriptions*'. Presumably, this distinction between neutral and action-descriptions pins down what we had in mind when saying, more vaguely, that physiologists 'describe their subjects in terms we could conceive to apply every bit as well to lab-rats or robots.' So let's agree to say that physiologists 'treat their subjects as meat-machines' just in case they describe the outward and internal behaviour of these organisms exclusively in neutral terms.

A further distinction, between different sorts of action-descriptions, will also be helpful: Both '*intention-sensitive*' and '*-insensitive*' descriptions imply that it makes sense to ask whether the behaviour in question was performed with an intention. But the former imply, in addition, that it actually was performed with

an intention. Our unease can then be put thus: 'Physiologists have no right to use action-, rather than neutral descriptions. And if they use the latter, their use of intention-sensitive descriptions is even more inappropriate than that of intention-insensitive terms.'

Let's get clearer on what the use of these different kinds of descriptions and, hence, the 'treatment' of subjects as 'meat-machines' and 'agents', respectively, comes down to in a pertinent experimental context: Consider discrimination experiments involving the by now familiar forced-choice task. Contrast two cases:

Case 1: The physiologists treat their subject as they treat human subjects in a reflex test: Before they start exposing him to the stimuli, they tell him, 'Now just sit back and relax. Don't get scared when your limbs start twitching.'

Case 2: Before they start exposing the subject to the stimuli, they explain to him what to do: 'This is the so-called "target-stimulus" we want you to keep track of. We'll show it to you a number of times, always accompanied by another stimulus that may be more or less similar to it. Press the right button whenever you see the target is to the right of the other stimulus, press the left button ... Whenever you are unsure which is which, just guess.'

In both cases, the physiologists expect responses. But in the second they expect their subject to be in direct control of those responses, here: the movements of his hands working the buttons, whereas they don't, in the first.

1 Contd.: The physiologists finish their instructions: 'Don't get scared when your limbs start twitching. Those are the entirely normal bodily reactions to the stimuli we are about to expose you to. Please don't spoil anything by trying to interfere.'

Here they are interested precisely in the behaviour that is not under the direct control of the subject: Precisely because they are interested in such behaviour, they ask the subject to refrain from exercising whatever indirect control he typically can exercise all the same, e.g., not to prevent his knee from jerking by blocking the movement with his arms. It is obvious that while faithful execution of these instructions might yield significant responses in a reflex-test, it would yield no relevant result in the present experimental set-up.

2 Contd.: The subject faithfully pushes his buttons as he has been told. At one point, his finger just twitches, pressing down the right button, and he reports it was just a twitch, and that he actually saw the target on the left. — Now contrast

Case 2-I: The experimenter says, 'Never mind', and records 'Right stimulus identified as target' or says 'Oh, all right' and records 'First right then left stimulus identified as target.'

Case 2-II: The experimenter records only 'Left stimulus identified as target' or nothing at all.

In both these cases, physiologists expect the subject to be in direct control of his button-pushing responses. But in case 2-I, they record those responses as stimulus-identifications regardless of whether they actually are under such control, whereas in case 2-II they count only those responses as identifications that are under the subject's direct control. I.e.: They use 'identify' as a neutral term in the former case, and as an action-description, in the latter.

II Contd.: At another point, the subject pushes first the left then the right button, and explains, 'That was a slip. I meant to press the right button.' — Now contrast

Case II-A: The experimenter records only the first or both button-pushings as stimulus-identification.

Case II-B: The experimenter records only the second button-pushing, or no identification at all.

In both cases, the subject is in direct control of his response movements, so that it makes sense to speak of them as being performed with some intention. In both cases, we assume, the experimenter is interested only in movements of which it makes sense to ask whether they are performed with some intention. But in case A he does not consider this question and counts the responses as identifications regardless of whether or not the subject responded in the way in which he intended to — or had any intention at all, for that matter. In case B, on the other hand, the experimenter does consider whether the subject meant to push the button he pushed and records that he identified such-and-such stimulus as the target only if the subject meant to point out that stimulus. I.e.: In both cases, the experimenter uses 'identify' as an action-description. But whereas he employs it as an intention-insensitive term in case A, its use in case B is intention-sensitive.

Now recall the definition: Two stimuli are indiscriminable for a subject iff he is unable to identify the test-stimulus at a significantly higher than random rate, over a series of trials. Whether (in)discriminability is a neutral or an action-notion thus depends upon the use of the verb 'identify'. As we just saw, this verb is used as a neutral description if the pertinent experimental practice is as described as 'case 2-I', and is used as an action-description, if that practice is in line with case 2-II. More specifically, it is used as an intention-insensitive description if experimenters' practice is as described as 'case II-A', and is used as an intention-sensitive description, if that practice corresponds to case II-B. So there is not only a neutral notion but also an action-concept of indiscriminability-for-a-subject. Indeed, this may even be an intention-sensitive concept. I.e.: Even in conducting discrimination experiments physiologists may fail to 'treat their subjects as meat-machines'.

At first sight, this will strike us as odd. This impression vanishes, once we see more clearly that 'indiscriminable' can be used *either* as an action- *or* as a neutral term, and that these two uses are, in a straightforward sense, 'on a par': Behaviour of a certain complexity *inevitably* confronts the experimenters with the question of which bits of it to regard as relevant. If, namely, the subject starts to correct his own initial responses in the way described as 'case 2', they have to make up their minds whether to take all responses into account, and if not, which to disregard. Whether they employ a neutral or an action-notion of discriminability then simply depends upon whether they give answer 2-I or 2-II to this question. Also if the subject corrects his initial responses in the way described as 'case II', the experimenters need to make up their mind as to which bits of behaviour to regard as relevant. Whether they employ an intention-sensitive or -insensitive notion of discriminability-for-a-subject then just turns on whether they opt for answer II-B or II-A. So all these uses are, in this sense, 'on a par' or 'analogous': They simply result from different answers to the question of which of different bits of behaviour of a certain complexity to take into account: the first response, the second, or both, or none?

To stress: In the context of the discrimination experiments we considered, behaviour of a certain complexity (viz., as in cases 2 and II) inevitably confronts the experimenter with the question of which of various responses to regard as relevant target-identifications. And whether he uses a neutral or an action-notion of discriminability, or an intention-sensitive or -insensitive action-notion, then simply depends upon which responses he decides to regard as relevant identifications, i.e., upon which of different bits of behaviour he decides to take into account. In this way, the different uses of '(in)discriminable' are 'on a par' or 'analogous'. And, in this way, also the two 'treatments' of subjects, as 'meat-machines' and as 'agents', respectively, can be regarded as 'analogous'.

Once we realise this, a remarkable change of attitude can often be observed: We no longer find it odd that 'indiscriminable' may be used *both* as a neutral *and* as an action-notion. And, more generally, we cease to find it either reprehensible or remarkable that physiologists do not merely use neutral, but also action-notions, indeed: even intention-sensitive such notions, in the description of the discrimination experiments we considered. We thus come to find it acceptable and pedestrian that physiologists may fail to 'treat their subjects as meat-machines', when testing for discriminability. Simultaneously, we cease to find it either remarkable or reprehensible that physiologists fail to 'treat their subjects as meat-machines' by testing for matching (rather than indiscriminability): We are no longer tempted either to object that they haven't got the right to do that, or to marvel at the fact that they brazenly do it. Thus, we are no longer uneasy about, or unduly impressed by, the practice of the physiologists we considered at the outset.

This way out of our predicament is no less odd than our way into it: After getting more clear on what we mean when venting our prejudice that 'physiologists treat their subjects as meat-machines', we considered in some detail what such 'treatment' and its opposite come down to in one particular experimental context. We then lost the inclination to react in the odd way outlined in the previous section, once we recognised that these 'treatments' are, in the given experimental context, 'on a par' or 'analogous', in the way explained. I.e.: Only the recognition of this analogy freed us from the grip of the prejudice that 'physiologists treat their subjects as meat-machines' that persisted as such an odd sense of unease or wonder, even once we had rationally disowned it. Borrowing some terms from Wittgenstein, one might speak of a 'therapy' that effects the relevant 'change of attitude' by overcoming the 'resistance of the will' that compelled us to react in the odd ways outlined.[5]

[5] This strategy to make sense of the later Wittgenstein's idea of 'philosophical therapy' requires explanation beyond the scope of this paper. For a detailed exposition see my 2000, ch.4.

7. Coda

Engaging in a bit of 'philosophical archaeology', we unearthed an objection to physiological explanations of experience many philosophers are shy about, even though they may admit that it, among others, moves them to find the very idea of such an explanation problematic. We refuted this objection, that 'physiologists can merely explain behaviour', and tried to dispel a peculiar sense of unease or wonder this left us with. To achieve this two-fold aim, we spelled out what the objector's intuitive contention ('Physiologists can merely explain behaviour') and a compelling slogan motivating it ('Physiologists treat their subjects as meat-machines') come down to when confronted with a concrete example taken from the pertinent (scientific) practice (sections 2-3) or come down to in a particular, pertinent (experimental) context (section 6), respectively. These means we put first (in section 4) to Moorean and then (in sections 5-6) to Wittgensteinian ends. Until a different philosophical objection starts to bother us, these two exercises jointly put us into a position to soberly appreciate what a particular explanation of experience actually offered by contemporary psychophysics can achieve: to appreciate that it explains nothing less than what it purports to explain; and that this achievement is not, in some peculiar way, more remarkable than equally clever physical explanations of other phenomena.

Hopefully, similar treatment of other philosophical objections to physiological explanations of experience will have the same success. The more this hope comes true, the less we will be inclined to wonder, 'How can facts about our physical nature explain why we have the experiences we have, in the circumstances in which we have them?' Then at least the second of our two starting questions will turn out to ask for an 'understanding' of 'how our physical nature could explain our subjective experience' that is to be gained not by constructing new philosophical theories (such as a new 'conception of our physical nature'), but by clarifying familiar philosophical questions and contentions in a particular way; namely, by relating them, in some detail, to work in the pertinent sciences. And by using the resulting clarifications, successively, to refute the philosophical contentions, and to pin down and dispel the disquieting sense of unease or wonder this may leave us with. With its modest but manifest success, this paper tried to recommend this two-step approach to your attention.[6]

[6] This paper was drafted during a stay at New York University made possible by a GAES grant and an invitation by Ned Block. For comments on a later version, I am indebted to Gordon Baker, Peter Hacker, Pierre Jacob, Bede Rundle, and the members of an audience at the École Normale Supérieure, Paris.

References:

Block, N. (1994), 'Consciousness', in: Samuel Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Oxford: Blackwell

Bock, G. (ed.) (1993), *Experimental and Theoretical Studies of Consciousness, Chichester & New York: Wiley*

Chalmers, D. (1996), *The Conscious Mind*, Oxford & New York: OUP

Clark, A. (1993), *Sensory Qualities*, Oxford: Clarendon Press

Comsweet, T. (1970), *Visual Perception*, London: Academic Press

Fischer, E. (2000), *Linguistic Creativity. Exercises in 'Philosophical Therapy'*, Dordrecht: Kluwer

Kirk, R. (1974), 'Sentience and Behaviour', *Mind* 83

Rundle, B. (1997), *Mind in Action*, Oxford: Clarendon

White, A. (1958), *G.E.Moore. A Critical Exposition*, Oxford: Blackwell

Wittgenstein, L. (1993), 'Philosophy', in his: *Philosophical Occasions 1912-1951*, Indianapolis: Hackett

Received: February 2001

Eugen Fischer
BPhil (Oxford), DPhil (Oxford)
Assistant Professor of Philosophy
Department für Philosophie
Ludwig-Maximilians-Universität
Geschwister-Scholl-Platz 1
D-80539 München, Germany