derten sich: Systematisierung statt System, Regulierung statt göttlicher Vorsehung, Vereinheitlichung statt Einheit. Schließlich führte die Berücksichtigung der Teleologie der Wissenschaft zu einer Betonung des pragmatischen Erfolgs der Wissenschaft als grundsätzlich deskriptivem Programm, das bewertet wird nach der Verwirklichung von Vorhersage und Kontrolle. Dies bedeutet, daß wir die Bestimmung der Einheit oder der Vereinheitlichung nicht als ein philosophisches Programm ansehen sollten, sondern als eine soziale Notwendigkeit, die erzielt wird durch die kooperative Arbeit der Wissenschaftler.

On the Disunity of Science or Why Psychology is not a Branch of Physics

MARTIN CARRIER

It is almost universally agreed that psychology cannot be adequately considered as a physical science but this consensus does not extend to the reasons advanced in support of this assessment. Rather, there are two sorts of pertinent arguments, the first dealing with the methodological aspects of psychology and the second with the nature of its subject matter. I will not delve into the methodological point but simply assert without further argument that there is no relevant difference between psychology and the physical sciences in this respect. Apart from some technical details, psychological laws and explanations exhibit the same basic features as physical ones, and the same holds for the methodological criteria applied in theory choice decisions. As regards the means and procedures psychology employs to approach its subject matter, it qualifies as a physical science.[1]

Things are different, however, with respect to the nature of the subject matter itself. The peculiarity that is usually associated with mental states is content. That is, mental states are said to be distinguished from physical states in that they express some thought or refer to some state in the external world. Mental states, but not physical ones, have the capacity to »mean« something. This is the traditional doctrine of intentionality, and what I try to do in this paper is to defend this doctrine. That is, I will argue that there is – at least for the time being – no adequate physical theory of the content of mental

1 For a more detailed elaboration of the methodological unity of science cf. M. Carrier and J. Mittelstrass, »The Unity of Science«, *International Studies in the Philosophy of Science* 4 (1990), pp. 1–15.

states. Whereas the thesis I wish to support has a rather traditional ring, the arguments I put forward in its favor do not, or so I hope.

There are at least two different options for an exploration of the relations between psychology and physics (or for that matter neurophysiology). In the first one the connections that exist or don't exist between concrete theories in these two disciplines or fields of research are examined. In the second option, the problem is addressed in a more abstract manner and the relations between psychological and physical theories in general are investigated. It is clear that for elucidating the relations between two comprehensive branches of science, the second option is preferable to the first. On the other hand, it can only be pursued if one has at hand a general scheme of the structure of a psychological theory or, in realistic terms, a general account of the specific traits of mental phenomena. The account of mental phenomena that I invoke for that purpose is the computational theory of the mind (developed in the framework of cognitive science). That is, I proceed on the assumption that this theory is correct, and I explore its bearing on the issue whether or not psychology is a physical science.

In doing so I will proceed in the following steps. I set the stage by, first, sketching the doctrine of intentionality and, second, by giving an outline of the computational theory of the mind. Then I turn to a discussion of one popular attempt to supplement the computational theory with an account of the content of mental states. Finally, I try to pinpoint the fundamental difficulty of this account (and similar ones). The degree of generality of the positions and arguments I deal with here has the uncomfortable consequence that my own discussion is of a rather abstract nature. This is as regrettable as it is unavoidable; the only thing I can do about it is to apologize for it. Sorry.

1. Intentionality or what's special about the mental

Mental states are often characterized by their capacity to refer to a certain thing or to an outside state of affairs. That is, they represent something and correspondingly possess content. If John believes that it is raining today his belief represents some outside circumstances, and if Jane wants to have her hair cut she wants to bring about a certain

situation. This reference of mental states to some external phenomenon has been called *intentionality*. Franz Brentano, who introduced the concept in 1874, considered intentionality to be the defining characteristic of mental states. All and only mental states are thought to possess the property of being »directed« to some state of affairs. Intentionality is another way of expressing the fact that mental states have content, i. e., that they are endowed with semantic attributes like truth-value or reference.

Brentano placed special emphasis on the fact that the object of a mental state need not exist in reality. Take the case, for example, that somebody has firm beliefs about unicorns or pink elephants. This does certainly not imply the actual existence of unicorns or pink elephants. Brentano coined the term ›intentional inexistence‹ to denote this peculiarity. It is a characteristic feature of a mental state that it may refer to a non-existent object.

It is of central importance for our problem that psychological explanations indeed make essential use of mental states of an intentional nature. They typically explain behavior by having recourse to mental states of a specific content. In order to explain the fact that John takes his umbrella when leaving his house, we ascribe to him the belief that it is raining (or that it will be raining). It is the belief content that is thought to bring about the behavior; beliefs are supposed to produce behavior by virtue of their content.

It is clear that the mere assumption of mental content does not explain anything; what we need in addition are laws in which this content enters. In fact, we are all familiar with such laws. A stock example is the so-called practical syllogism that has the general form:

> For all goals and beliefs: If A wishes that *b* be the case and moreover believes that *a* is an appropriate means for achieving *b*, A sets out to realize *a*.

Laws of that sort are called *folk psychological* laws. It is characteristic of folk psychological laws that they make essential use of intentional states; they generalize over such states. After all, the preliminary clause of the law says that the law is supposed to extend to all goals and beliefs. Intentional states are the quantified variables of the law and thus constitute its universe of discourse.

It deserves emphasis at this juncture that the laws of empirical or scientific psychology also display this characteristic feature. That is, they equally generalize over intentional states. True, the laws of scientific psychology are more sophisticated than those of folk psychology and they exceed by far the explanatory power of the latter. But while there is a difference in substance between the two, their conceptual type is much the same. Explanations in scientific psychology likewise rely on the content of mental states; they rely on the capacity of a person to distinguish between relevantly different aspects of a situation and to evaluate these aspects. Scientific psychology assumes, in other words, that an outside state of affairs is cognitively represented by a system of beliefs and that the actions of a person are guided by that system (along with a system of preferences and goals). This means that the cognitive apparatus that scientific psychology ascribes to a person in order to account for this person's behavior is of the same general nature as the corresponding apparatus used in folk psychology.

Now we are in a position to formulate our general problem. On the one hand, it appears that science has some use for content; on the other hand, content is obviously an extremely elusive entity. So we are left with the task of clarifying what sorts of things contents actually are and how on earth they manage to perform the feat of producing behavior. In what sense can mental content be the effect of some sensory input or the cause of some behavioral output?

This job appears hard enough at the outset, but it will appear even harder if one realizes the intricate relation between mental content and physical states of affairs. The point is that both cross-classify one another in a peculiar fashion. Take, for example, the problem of determining the type identity of a belief state. That is, we are asking under what physical circumstances two mental tokens (e.g., two belief states realized in different persons or in one person at different points in time) are of the same type (i. e., are belief states of the same content).

Let's assume that somebody holds the belief that it is raining. This simple belief state is associated with physical indicators (i. e., with sensory stimuli, behavioral responses, and neurophysiological processes) in an extremely complicated manner. Let's begin by considering its possible sources, i. e., the physical situations that may precede it.

Someone may come to the belief that it is or will be raining by watching the weather forecast on TV or by listening to the radio. He may read it in a newspaper or infer it from some characteristic pains he feels in his limbs. The physical difference between the possible sources of the belief implies that the neurophysiological processes involved in the formation of the belief are also different. It certainly makes a neurophysiological difference whether the belief is formed via a stimulation of the optical center of the brain or by means of its acoustic center.

The same holds analogously for the physical consequences of a belief state. That is, the same belief state may lead to different actions. The belief that it's raining outside may occasion somebody to reach out for his umbrella, to take the car or the bus or to stay at home altogether. It is clear that these actions are realized by means of different muscular innervations and thus different neurophysiological processes.

This leads to the following interpretation: The physical (i. e., sensory, behavioral and neurophysiological) phenomena associated with one and the same kind of belief are extremely variegated. One psychological state is linked to a plethora of physical states, and the only common ground between these physical states is that they are all tied to the same psychological state. They have no physical characteristic in common; they cannot be determined or demarcated against other physical states by relying on physical properties. The set of psychologically equivalent physical states is constituted – from the physical point of view – by a wild disjunction of unrelated phenomena. There is no physical law that collects them into a class. Their sole connection is that they end up in or start off from the same belief state. And the crucial question is, then, what is the physical justification for considering this state to be the *same* belief state.

An analogous difficulty of physically reconstructing the principles of mental token typing is encountered in the converse situation. Up to now it has been argued that there is a one-many relation between psychological states and physical states; there exists, however, a many-one relation in addition. That is, one physical situation can be the cause or the effect of various different psychological events. Let's consider two TV viewers watching the weather forecast and learning that it will rain tomorrow. The first may be convinced of the reliability

of the forecast whereas the second may be extremely skeptical about its trustworthiness. In that case, consequently, one physical state of affairs gives rise to two quite different belief states, namely, the conviction that it will rain tomorrow in the first person, and a state of agnosticism in the second one. Conversely, one and the same action may originate from different belief states. Different people (or one person at different points in time) may do the same thing for different reasons.

So all in all there is a many-many relation between physical and psychological states. Different physical stimuli may lead to the same belief state and the same physical situation may generate different belief states. And so the question arises what ties together the physically dissimilar stimuli and what separates the physically similar ones. How can we physically characterize the fact that two persons possess the content-identical belief that it's raining while neither the antecedent conditions nor the consequent actions coincide?

Systematically speaking the problem can be put as follows. Every corpus of laws introduces a taxonomy into its universe of discourse. By associating certain properties with certain objects in a lawful manner, it binds these objects together; it establishes a link between these objects by regarding them as instances of the same law. That is, laws induce a system of token typing in their domain of application, and these induced types are called *natural kinds*. Take, for example, the law: All ravens are black. This law creates the natural kind »raven« by quantifying over ravens. Ravens constitute a natural kind because there is a law that applies to them by virtue of their property of being ravens. By contrast, think of the generalization: All things on my desk are in a terribly disordered state. Since this is not a law but merely an accidental generalization (however true), the variables over which it quantifies (i. e., the things on my desk) do not constitute a natural kind. Though they certainly have various properties in common (after all, they are all physical objects), the common ground among them does not arise from the fact that they are all instances of the above-mentioned generalization. So while there are laws about ravens (or electrons, planetary systems or the like), there are no laws about the things assembled on my desk or the objects located within a distance

of three kilometers around the university of Konstanz. Accordingly, the former classes constitute natural kinds and the latter don't.[2]

With respect to psychology it emerges from the discussion above that belief states (that is, mental states with a certain content) are members of a natural kind. Psychology generalizes over belief states and thus transforms them into a natural kind. True, this holds in the first place only for the mental type »belief state« (in contrast to other mental types such as goals), not for the content of belief states. It is transparent from the above-mentioned law of practical syllogism, however, that sameness or difference in content is essential for the application of laws that make use of belief states (or intentional states in general). Namely, such laws invoke content to identify belief states, as it can be seen from the procedure of using the same variable to denote mental states of the same content.

The problem of intentionality now comes down to the following question: In which way and by virtue of which properties can we relate physical natural kinds to cognitive natural kinds? And this problem is made non-trivial (or rather extremely hard to solve) by the cross-classification of both types of natural kinds. To make psychology a branch of physics requires the solution to this problem; that is, it requires that we derive the principles of psychological token typing (i. e., psychological natural kinds) from physical natural kinds. I will discuss the possible approach to a solution to that problem in two steps. In the first step I will present the basics of the computational theory of the mind, and in the second one I will address the problem of psychosemantics proper.

2 The concept of natural kinds in the above-mentioned sense is due to Fodor; cf. J. A. Fodor, »Special Sciences (or: The Disunity of Science as a Working Hypothesis)«, *Synthese* 28 (1974), pp. 101–102. The cross-classification of physical and psychological natural kinds and the differences in taxonomy they induce among objects is stressed by J. A. Fodor, »Special Sciences«, pp. 103–107; Z. W. Pylyshyn, *Computation and Cognition: Toward a Foundation for Cognitive Science* (Cambridge, Mass., 1984), pp. 7–12. Cf. also M. Carrier and J. Mittelstrass, *Geist, Gehirn, Verhalten: Das Leib-Seele-Problem und die Philosophie der Psychologie* (Berlin/New York, 1989), pp. 70–75, pp. 205–206.

2. *The computational theory or how the mind works*

The computational theory of the mind is a partial answer to the problem of how content manages to produce overt behavior. The gist of that answer is that in fact content does not produce anything. After all, it's hard to imagine how the content of a belief could influence the activation state of a synapse. The point is, however, that content is tied to the physical aspects of mental states in such a way that the causal connections between these states respect the semantic constraints imposed by their content.

This theory of the mental machinery is modeled on the functioning of computers. Computers indeed operate successfully with semantically interpreted magnitudes. They calculate optimal profiles for airplane wings or simulate complex oscillations of large systems; they transform, accordingly, interpreted input data into interpreted output data. But they do so not by having recourse to the semantic attributes of these data; rather, they translate these data into a formal, internal language. Computers generate strings of uninterpreted symbols from the input data and operate on these strings following formal rules programmed into them.

Take, for example, the addition $9 + 5 = 14$. Here the symbols have content, they are interpreted as numbers. For carrying out such an addition a computer first transforms the numbers into strings of formal symbols. In particular, it assigns the following strings to the numbers: $9 \rightarrow$ xoox, $5 \rightarrow$ xox. This is in fact the binary encoding of the decimal numbers but in our context we can treat these strings simply as uninterpreted tokens. Second, the computer applies a rule to the effect: If (say) register 1 is in a state described by the formal string xoox, and register 2 is in a state characterized by xox, then change the state of register 3 into a state formally denoted by xxxo. And the latter string is, third, retranslated into the decimal number 14.[3]

3 It is clear that the rule invoked here has an extremely narrow domain of application and is for that reason rather unrealistic. Actual rules are constructed such that they are applicable to a wider range of cases. But the principle of operation remains the same even in more complex examples.

A program of this sort successfully creates the impression that it can sensibly operate with semantically interpreted magnitudes. In fact, however, the program has obviously no access whatsoever to the semantic interpretation. The program feigns to operate with numbers according to the rules of addition whereas in fact it only recognizes differently shaped symbols and treats them according to rules that are by no means content-sensitive. While the symbols have lost their meaning during the translation into the internal, formal language, the program works as if the symbols were still interpreted and as if the rules had access to this interpretation. In order for this formal mimikry to be possible, the whole process has to satisfy the so-called *formality condition*: The program and the translation procedure have to be designed such that all relevant differences in semantic content are reflected in formal differences between the associated machine states.

The central step of the computational theory lies in transferring this model to the human mind. To think is literally to run through a computer program. This means that a chain of neurophysiological states whose evolution in time is governed by causal-physical laws can also be regarded as a sequence of strings of symbols that is generated by certain formal transformation rules. These rules are in turn of such a nature that they respect the logical and semantic relations between the content that can be associated with these formal states. If the formality condition holds in the brain, human reasoning can be described as content-based inference and at the same time as formal symbol manipulations and causal state transitions. The formality condition guarantees that processes that are blind to content may well respect content-based restrictions.[4]

4 For this sketch of the computational theory cf. J. A. Fodor, *The Language of Thought* (New York, 1975), p. 32, pp. 66–67, pp. 73–74; J. A. Fodor, *Representations: Philosophical Essays on the Foundations of Cognitive Science* (Cambridge, Mass., 1981), pp. 226–227, pp. 240–241; Z. W. Pylyshyn, *Computation and Cognition*, pp. 26–40, pp. 59–61; K. Sayre, »Cognitive Science and the Problem of Semantic Content«, *Synthese* 70 (1987), pp. 247–251. Cf. also M. Carrier and J. Mittelstrass, *Geist, Gehirn, Verhalten*, pp. 207–210. Taking the functioning of a computer as a model of mental activity implies, incidentally, that for the computational theory artificial intelligence really is artificial intelligence.

This tri-level structure, i. e., the hierarchy of the causal-physical, the formal-syntactic, and the content-based, semantic level, lies at the heart of the computational theory and of cognitive science in general. If this model is supposed to adequately capture the machinery of the mind, it must induce the right relations between physical states and mental states; that is, the model must entail the consequence that a many-many relation exists between physical and mental types. This is indeed the case. The computational theory in fact implies the existence of a many-many relation between physical and syntactic types of state as well as between syntactic and semantic types of state. In order to realize this we must first recall an important peculiarity, namely, the multiple interpretability of formal structures.

One and the same formal algorithm can be applied to various, different domains. Take for example the following abstract equation:

$$\alpha_1 \frac{d^2u}{dt^2} + \alpha_2 \frac{du}{dt} + \alpha_3 u = 0$$

This equation describes damped oscillations in a general or formal manner and can be applied to several different systems by interpreting the abstract variables in a different fashion. Consider for instance the following interpretation: $u \rightarrow$ elongation of a spring, $\alpha_1 \rightarrow$ mass suspended from the spring, $\alpha_2 \rightarrow$ viscosity of the medium in which the oscillation takes place, $\alpha_3 \rightarrow$ elastic constant of the spring. In that interpretation the equation deals with the oscillations of a coil spring in a resistant medium such as air. But this is by no means the only possible interpretation. The abstract variables may as well be interpreted in a quite different fashion. Consider the following assignment: $u \rightarrow$ current intensity, $\alpha_1 \rightarrow$ self-inductance of an electric circuit, $\alpha_2 \rightarrow$ electric resistance, $\alpha_3 \rightarrow$ inverse capacity. In this interpretation the equation treats the oscillations of the current intensity of an electrical circuit equipped with a capacitance, a self-inductance and a resistance. The same equation thus models two situations which are quite distinct with respect to content; the same equation has two quite distinct semantic interpretations.

Let me illustrate this important peculiarity by another example. I start by briefly outlining the equation for the production rate of a

substance in an autocatalytical chemical reaction and afterwards apply this equation to two further domains which intuitively have nothing in common with chemical reactions. Assume that the substance A is produced by an autocatalytical reaction of the form: $A + B \rightarrow 2A$, and assume furthermore that B is supplied from outside with a constant rate and that A decays with a likewise constant rate. In that case the net production rate of A molecules is the number of A molecules generated minus the number of A molecules lost. Because of the autocatalytical form of the production process the gain of A is proportional to the number n of A molecules already present ($dn/dt \sim \alpha_1 n$), and its loss is also proportional to n ($dn/dt \sim -k_1 n$). Moreover, the reaction consumes B molecules and thus leads to a reduction of the number N of B molecules that is proportional to the reaction rate and accordingly also proportional to n ($\Delta N = -\alpha_2 n$). Because of the limited supply of B molecules, this effect reduces the production of A molecules. Plugging in these expressions into the basic »gain-minus-loss« approach (and renaming the constants where appropriate) gives rise to the following overall equation for the temporal development of the number of A molecules:

$$\frac{dn}{dt} = -k_1 n - k_2 n^2$$

This is a nonlinear differential equation whose solutions approach two stable states (depending on the value of the constants).

The point is that the very same equation likewise describes the number of photons in a cavity with induced photon emission, i. e., a laser system, and the temporal evolution of population size in a herbivorous species in a bounded region. In the first case, the gain in the number of photons is due to photon-stimulated emission of photons from excited atoms, the loss stems from the escape of photons through the endfaces of the laser, and the reduction term expresses the decreasing number of excited atoms which is a result of the emission process itself. The basic traits of this laser model obviously coincide with those of the autocatalytical reaction, and accordingly the fundamental equations of both processes coincide as well.

The same holds with respect to the ecological case. In that case, n refers to the number of the members of a zoological species living in

an area with limited food supply. The birth rate as well as the death rate of the population is proportional to the number of living animals, and the reduction term appears as a consequence of the depletion of the food resources. So in this application the equation models the temporal development of a population of herbivores.[5]

These two examples demonstrate the multiple applicability of formal algorithms. The same abstract equation holds in intuitively quite different cases. This multiple applicability is due to the fact that the variables in this equation refer to different things in either case; the variables respectively mean different things. This important feature can be expressed such that formal structures do not uniquely determine their own interpretation; they only determine a whole set of interpretations which differ in meaning but display the same formal aspects. Formal structures determine meaning only up to isomorphism.[6] It is to be noted in addition that this multiple interpretability is not restricted by the requirement that the several instantiations of the same formal structure all actually be realized. It is perfectly legitimate, by contrast, to interpret a formal structure in such a way that it refers to a fictitious state of affairs. It is precisely the capacity to refer to non-existent situations that constitutes Brentano's intentional inexistence.

Finally, there is a famous case in the philosophy of psychology that can be used to shed some additional light on the point at issue. Namely, the problem of the inverted spectrum that was first pointed out by Locke. Locke imagines that the same physical signal might give rise to different color perceptions in different men. A violet might produce the same color quality in one man's mind that a marigold produces in another one's.[7] This amounts to a systematic permutation of sensory qualities, or for that matter of mental content, associated with the

5 For this example cf. H. Haken, *Synergetics: An Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry and Biology* (Berlin/Heidelberg/New York, 1978), pp. 127–128, pp. 266–267, pp. 293–294.

6 This peculiarity also underlies Searle's well-known »Chinese Room Argument«; cf. J. R. Searle, »Minds, Brains, and Programs«, *The Behavioral and Brain Sciences* 3 (1980), pp. 417–424.

7 Cf. J. Locke, *An Essay Concerning Human Understanding* (1690), ed. P. H. Nidditch (Oxford, 1975), p. 389 (II. XXXII.15).

same signal. The external property that brings forth the quality or content »blue« in a person A, brings about a quality or content in a person B that A would call »yellow.« Systematically speaking, the inverted spectrum is a map among color qualities that preserves the relations between these qualities. For the inverted viewer green is still located between blue and yellow, but the mental content associated with these color shades is exchanged.

A syntactic approach is obviously unable to capture such a content inversion. After all, the sequence of formal states involved in formation of beliefs about colors remains unaffected by the inversion. In the normal-eyed viewer the content »blue« plays the same formal or functional role that the content »yellow« plays in the inverted one. The content inversion has no impact on the formal machinery of the mind. In the case of the inverted spectrum we can attach different content to the same formal variable without thereby altering in any sense the formal description of the state transitions constituting the mental activity. The normal-eyed and the inverted viewer have the same syntactic description. This shows once more that formal structures are unable to differentiate between isomorphic interpretations.

On the basis of this insight we can now easily recognize that the computational theory of the mind indeed implies the existence of a many-many relation between physical and mental types. Let's assume for the sake of simplicity that there exists a straightforward correspondence between the input data and the physical state of the receptor system. In that case the cross-classification of types of external states and cognitive types assumes the form of a cross-classification between internal physical types (i. e., types of machine or brain states) and cognitive types. A many-many relation of this sort indeed holds. More specifically, it even holds for both interlevel connections, i. e., between the semantic and the syntactic level as well as between the syntactic and the physical one.

As the foregoing discussion has shown, one and the same formal algorithm may have various, different applications or interpretations. This implies that one and the same sequence of formal states may represent a variety of states of affairs. The converse is equally true: One and the same intuitively understood system can be modeled by differently structured equations. After all, one is always free to intro-

duce some superfluous complexities into a given equation and to transform it in this way into a structurally different one. Moreover, the discussions on inductive simplicity and the Duhem-Quine thesis contain a lot of additional evidence for the option to cope with a given system or problem-situation in various, theoretically distinct ways. These results can be immediately transferred to the relation between semantics and syntax in machines or brains. The same sequence of semantically interpreted states can be described by different formal algorithms. This leads to the conclusion that there is indeed a many-many relation between semantic interpretations and syntactic structures.

The same holds with respect to the relation between syntax and physics. It is clear that the same formal structure – i. e., the transformation of certain initial strings of symbols into output strings – can be realized by means of several distinct physical systems. That is, the same computer program (as described on the level of formal rules) can be implemented on different material computer systems. So there is a one-many relation between the syntactic and the physical level.

Conversely, any such physically described computer state can represent different formal structures by means of a change in the association between physical and syntactic types. A syntactic type is usually associated with a class of physically distinct states; that is, for instance, we abstract from voltage fluctuations and correspondingly associate the same symbol with a whole class of neighboring voltage states. By changing the range of physical differences we are willing to neglect we can generate different syntactic descriptions of the same physical state. Deliberate decreasing or enlarging of the syntactic equivalence classes generates different syntactic descriptions of the same physical system. Another option consists of associating changes of physical parameters (instead of their absolute values) with a syntactic symbol – as it is done, for instance, in the digital code of a compact disc. Accordingly, there is a many-one relation between the syntactic and the physical level in addition, and this leads to the overall conclusion that a many-many relation between syntax and physics is present as well.

What is to be concluded from all that regarding the problem of intentionality? The discussion makes it clear, first, that and how physical systems can create the impression of possessing intentional states

while in fact they possess nothing even remotely similar to such states. It shows, second, that whereas the computational theory gives an account of mental operations (and represents in fact the only worked-out account in that field), it has nothing to say on how mental states acquire content. After all, we believe that the intentionality of mental states is real and not deceptive, and this assessment is underpinned by the fact that content-based psychological laws provide the best available explanation of human behavior. In contrast to mental states, however, intentionality is conferred to computer states solely through human interpretation; as regards the latter, intentionality resides in the eye of the beholder. It should be noted, third, that the inability of the computational theory to accomodate mental content is a matter of principle. The theory cannot be amended such that in the end it is able to cope with content. For the theory turns on the multiple interpretability of formal algorithms; it is this feature (among others) that furnishes the right sort of relations between mental and physical types. On the other hand, it is this very same feature that rules out that the computational theory contains an account of mental content. So it is the same trait that makes the theory work as a model of mental operations and the nature of human intelligence and that makes it at the same time unfit as a model of intentional states. This can be summarized such that a theory of psychodynamics does not entail, but rather has to be supplemented with, a theory of psychosemantics.

3. Psychosemantics or how to cope with mental content

The problem of psychosemantics now comes down to the task of supplying a formal structure with content in a proper and adequate way. That is, we must explain how content, or for that matter the type identity of belief states, can be assigned to a formal structure by exclusive recourse to physical (i. e., non-semantic) means. This outlines the project of a naturalized psychosemantics. A promising strategy for implementing this project seems to be the following: The semantic ambiguity of formal structures stems from the fact that they are purely formal. So let's endow them with content by attaching referents to some of the symbols occurring in them. After all, we have already

successfully applied precisely this recipe for removing semantic ambiguity. Recall the examples of the damped oscillations or the autocatalytical processes discussed above. A definite meaning has been conferred to the abstract variables by associating referents with them. What we did was to stipulate that, e.g., the variable *u* is to refer to the elongation of a spring or, in the alternative interpretation, to current intensity. That little is sufficient to generate content from formal systems. And what has been successful in these physical examples should also work for mental states, or so it seems. So let's examine how such an approach fares in the mental area.

All approaches to a naturalized psychosemantics rely in some way or other on this procedure. They attempt to transform an abstract structure of mental tokens into a system of beliefs by imparting reference to at least some parts of that structure. For reasons of perspicuity I will pick but one example from the class of like-inspired attempts, namely, Fred Dretske's information-theoretic model of cognitive content. The discussion of that model has to be very brief and sketchy, and so I will pass directly to my essential point. That is, I will focus on what I take to be its crucial shortcoming. This concentration on the model's vices might lead to an overly negative appraisal, and so I should state in advance that it also possesses important virtues and that only limited space precludes dealing with the latter more thoroughly.

The central tenet of the information-theoretic account is that all representation is correlation. A physical system represents another physical system if their respective states are lawfully connected. More specifically, suppose that the state of a system S_2 is deterministically related to the state of a system S_1. Such a situation occurs if the S_1-state causally produces the S_2-state and if no other possible causes of this S_2-state are realized. In that case the S_2-state indicates the presence of the corresponding S_1-state, and this can be expressed such that the former gives information about the latter. So there is a relation of reference or aboutness contained in such a causal connection, and this relation constitutes the most fundamental feature of intentionality.

It is clear that this basic mechanism has to be supplemented with some auxiliary procedures. For up to this point, intentionality has lost its distinctive property of characterizing mental phenomena. Thermometers, galvanometers and the like may well possess intentional states;

their internal states are connected in a lawful fashion to some outside state of affairs and thus represent it. Accordingly, intentionality appears to pervade the whole of nature. In order to avoid this unwelcome consequence Dretske adds two further requirements that are supposed to distinguish truly cognitive systems. Such systems must possess the capacity of *aspect separation* (my term) and »*digitalization.*«

The first requirement is based on the observation that non-cognitive, representational systems don't represent just one external state but rather a large cluster of such states. Think, for instance, of a galvanometer that measures current intensity. The readings of that instrument do not represent that quantity alone; in addition, they represent voltage states (by virtue of Ohm's law). One and the same state of the instrument thus represents several intuitively distinct physical quantities, and this shows that this state does not carry a definite or unique piece of information.

Cognitive systems, by contrast, are required to be able to separate the different aspects contained in a given physical signal. That is, they must possess the capacity to represent the same external state in different fashions. Whereas in nature a given state of current intensity always occurs together with a certain voltage state, a cognitive system has to be capable of representing one of them without at the same time representing the other.

The second characteristic of cognitive systems is their capacity of digitalization. Digitalized representation means that not all states that are different in the system S_1 are represented as different states in the system S_2. The representing system S_2 collects distinct states of S_1 into equivalence classes; i. e., it represents them as the same states. Think of the cognitive representation of a table as an example. This concept embraces a large number of physically distinct and heterogeneous entities. Tables may be made of wood or of some other material, they may be painted in different colors or located in various places. In the cognitive representation of a table as a table we abstract from all these peculiarities and collect the different tokens into one equivalence class; we put the common label »table« on them.[8]

8 This reconstruction of the information-theoretic account is based on F. I. Dretske, »The Intentionality of Cognitive States«, *Midwest Studies in*

On Dretske's view it is these two procedures, namely, sorting out components of information and stripping off such components, that are characteristic of cognitive representation. The introduction of these procedures serves the purpose of reproducing theoretically the many-many relations between physical and mental types. Aspect separation is intended to make sure that the same physical situation may be represented differently, and digitalization is conversely supposed to establish the possibility that different physical situations are represented in a like fashion. In this way the model is thought to accomodate the cross-classification of physical and mental natural kinds. The effect of the combined application of both procedures is thus a reshuffling of physical signals according to their cognitive significance. Physically alike signals may be dissociated and put into different cognitive equivalence classes whereas physically distinct signals may be associated and placed in the same cognitive equivalence class.

The central question that emerges at this juncture is: What are the principles that guide this reshuffling process? A naturalized psychosemantics requires that the principles governing the formation of cognitive equivalence classes be specified in physical, i. e., non-intentional and non-semantic, terms. Cognitive types must be derived from physical types by purely non-semantic means. But Dretske's account contains no clue whatsoever how this job is to be performed. We need a criterion that explains how and by virtue of which non-semantic properties two physically distinct signals can »mean« the same thing or vice versa. Without such a criterion the whole reshuffling procedure is merely an arbitrary combinatorial exercise. For a naturalized psychosemantics it is not sufficient to reproduce the right sort of relations between mental and physical types; we need some rationale for establishing these relations precisely. In the absence of such a rationale, cognitive equivalence classes can only be framed by relying either on arbitrary chance mechanisms or on semantic intuition. And since the first option clearly makes no sense, Dretske's account of psychosemantics is in the end founded on meaning in disguise.

Philosophy 5 (1980), pp. 281–294; F. I. Dretske, »Précis of *Knowledge and the Flow of Information*«, *The Behavioral and Brain Sciences* 6 (1983), pp. 55–90.

This essential flaw is not confined to Dretske's model; all other versions of a correlational psychosemantics are beset with the same or similar difficulties. None of them succeeds in elaborating principles of mental token typing that solely rely on non-semantic procedures. Instead, mental token typing is simply read off from linguistic experience. The conclusion is that we are at a loss – at least at present – to explain by virtue of which non-semantic principles content-based equivalence classes are to be extracted from physical natural kinds. Content cannot be captured by purely physical criteria, or, in other words, cognitive representation is a very special and most peculiar relation. This is why psychology is not a branch of physics.

It is to be stressed, on the other hand, that this result essentially depends upon the correctness of the computational theory of the mind. If this theory should turn out to be completely off the mark, the whole problem has to be reconsidered. The validity of the analysis depends, second, on the hypothesis that mental states are of an intrinsically intentional nature. It is to be noted, third, that important aspects of the working of the mind can in fact be accounted for. Foremost among these aspects is the nature of mental operations. The explanation of what the mind actually (or possibly) does is certainly a primary achievement – regardless of the fact that some characteristics of the objects of these operations still defy physical analysis.

There is one problem left that should be briefly dealt with at the end. With hindsight one might entertain some doubts with respect to the above assertion that the computational theory provides an adequate account of mental operations. The reason is that the relation between physical types of states and syntactic ones is as intricate as the corresponding relation between syntactic types of states and semantic ones. We have a many-many relation on either level. So we can draw an analogous conclusion: The principles of syntactic token typing are not derivable from the physical theory of the underlying mechanism.

There is in fact no physical justification for representing two physically distinct computer states by the same symbol. A formal representation of a given physical system always amounts to collecting physically distinct states into a syntactic equivalence class. Conversely, the same physical state can be described by several syntactic structures. This feature gives rise to the existence of a many-many relation between

syntax and physics that I tried to outline above. In view of the importance computational theory attaches to the syntactic level, this seemingly arbitrary character of syntactic token typing may induce some reservations concerning the relevance of syntactic operations for the functioning of the mind.

These misgivings are, however, unfounded. True, we haven't a physical theory of syntactic token typing but we don't need one either. Every syntactic token typing that is compatible with the sequence of machine states run through is all right; every token typing that meets the physical constraints is legitimate. Some may be more elegant and some more clumsy but apart from these pragmatic aspects they have all equal status. The reason is that the syntactic description is of purely instrumental value. What matters is the right connection of semantic states that is in turn brought about by the right connection of physical states. The syntactic level is only introduced for descriptive purposes.

By contrast, the semantic interpretation is of intrinsic and not of merely instrumental importance. Mental representations really exist in cognitive systems, or so most of us believe. It is this feature that makes the project of developing a naturalized psychosemantics at all interesting and worthwhile. The project of a »physicalized psychosyntax,« on the other hand, is of no use whatsoever. True, both projects fail for the same reasons, but as regards the second one nobody needs to feel bothered about that. So it is justified to uphold the assessment that the computational theory gives an adequate account of mental operations. It is not the nature of intelligence that still constitutes a mystery but rather the nature of intentionality.

Summary

Untersucht wird das Verhältnis zwischen den Gegenstandsbereichen der Physik und der Psychologie. Dabei wird von den beiden Voraussetzungen ausgegangen, daß (1) Intentionalität kennzeichnendes Merkmal geistiger Phänomene ist und daß (2) die sogenannte Rechnertheorie des Geistes eine im Kern zutreffende Charakterisierung mentaler Operationen bereitstellt. Daraus ergibt sich, daß der Gegenstandsbereich der Psychologie nicht mit physikalischen Mitteln erfaßt werden kann. Der Grund ist die mehrfache inhaltliche Interpretierbarkeit formaler Strukturen, welche zur Folge hat, daß die natürlichen Arten der Psychologie nicht aus den natürlichen Arten der Physik abgeleitet werden können. Die Grundsätze der Bildung psychologischer Arten können auf physikalischer Basis lediglich beschrieben, nicht aber erklärt werden.