"How does it work?" vs. "What are the laws?"
Two conceptions   of psychological   explanation.

Robert Cummins

University of California-Davis

1. In the beginning

In the beginning, there was the DN (Deductive Nomological) model of explanation, articulated by Hempel and Oppenheim (1948). According to DN, scientific explanation is subsumption under natural law. Individual events are explained by deducing them from laws together with initial conditions (or boundary conditions), and laws are explained by deriving them from other more fundamental laws, as, for example, the simple pendulum law is derived from Newton's laws of motion.

It is well-known that DN is vulnerable to a wide variety of counter-examples (e.g., Kim, 1962; Salmon 1998). As a result, DN is not widely defended. But it is, I think, still widely believed that scientific explanation is subsumption under law. This is something of a scandal: Given DN's miserable track record in spite of spirited defense by many ingenious believers, one is led to ask why so many cleave so faithfully to a doctrine that has proved so indefensible?

There are two factors that work to keep DN in place. First, there is the fact that every experimental paper one picks up involves the explanation of some data by appeal to some hypothesis or other. It is tempting to conclude that philosophers' continued failure to articulate this practice in some defensible way is a point against philosophers, not against DN. And second, there is the fact that there is no widely understood and compelling alternative to DN on the

market. If cognitive psychology has taught us anything, it is that no one willingly gives up a well-worn idea without having something to put in its place. I propose to examine these two factors in turn.

2. <u>Two Pitfalls</u>

In psychology, DN gets a spurious plausibility from the fact that data are routinely said to be explained or accounted for by some hypothesis or other. But this is likely to be misleading in at least two ways.

First, when psychologists talk about explaining or accounting for some percentage of the variance, the "hypothesis" in question is that that the experimental treatment will have some real effect. One is looking to reject the null hypothesis in favor of its complement, namely the hypothesis that whatever differences there are between the treatment group and the control group are not due to chance (random variation). But this sort of hypothesis isn't a law or anything like a law. The word "hypothesis" as it is used in statistical analysis, and the word "hypothesis" as it is used to refer to a conjectured theory or law, are little more than homonyms: They share the element of conjecture and little else. While there is nothing wrong with either use of the word, it is well in contexts like the present one to keep the two concepts distinct. With this in mind, I will use "law" to refer to hypotheses in the second sense.

The second way in which talk of explanation in the context of the statistical analysis of data is likely to be misleading is that, even though experimenters sometimes are attempting to test a theory or an hypothesis in the second sense (i.e., a proposed law or regularity), this is an exercise in confirmation, not explanation. We say that a law or theory accounts for or explains the data, but this simply means that the data confirm the law or theory. When a law is confirmed by some data set, this is evidence that the law

<u>describes</u> the data (to some reasonable approximation or other). The now classic illustration of this is Balmer's formula (Hempel, 1966, pp. 37ff).

$$\lambda = 3645.6 \frac{n^2}{n^2 - 4}$$

This formula specifies the wavelengths of the emission spectrum of hydrogen. Finding spectral lines in the places predicted by the formula confirms the law, but no one thinks the law explains why the lines are where they are.

Defenders of DN concede that Balmer's formula and similar cases are cases in which subsumption under law is not explanatory. They then take their task to be formulating a criterion that will distinguish cases like Balmer's formula from genuinely explanatory laws. There is wide consensus, however, that this has not been done successfully, and the suspicion grows that it <u>cannot</u> be done successfully. I think we should take seriously the possibility that it cannot be done because there isn't any difference: No laws are explanatory in the sense required by DN. Laws simply tell us what happens; they do not tell us why or how. Moli ère, satirizing scholastic appeals to occult properties and "virtues", tweaks the doctors of his time for explaining that opium puts people to sleep because it has a dormitival virtue. But isn't this just what subsumption under law always amounts to? Does the Law of Effect explain why giving a pigeon Pigeon Chow whenever it pecks a key increases the rate of key pecking? Or does it just restate the phenomenon in more general terms? Surely the correct moral to draw here is that the Law of Effect is an explanandum, not an explanans.

In science, when a law is thought of as an explanandum, it is called an "effect." Einstein received his Nobel Prize, not for his work on relativity, but for his explanation of the photo-electric effect. In psychology, such laws as there are are almost always conceived of, and even called, effects. We have the

Garcia effect (Garcia, 1966), the spacing effect (Madigan, 1969), the McGurk effect (MacDonald and McGurk, 1978), and many, many more. Each of these is a fairly well confirmed law or regularity (or set of them). But no one thinks that the McGurk effect explains the data it subsumes. No one not in the grip of the DN model would suppose that one could explain why someone hears a consonant like the speaking mouth appears to make by appeal to the McGurk effect. That just is the McGurk effect.

The mistaken idea that accounting for data by subsuming it under la w is explanation is also fostered by a confusion between explanation and prediction.[1] A law that predicts a certain data point or data set is said to explain it. But prediction and explanation are separable in ways that DN cannot accommodate. It is possible to understand how a mechanism works, and hence to be in a position to explain its behavior and capacities --the effects it exhibits -- without being able to predict or control its behavior. This is true generally of stochastic or chaotic systems. It is also true of systems whose relevant initial states are unknowable or simply unknown. In possession of a machine table for a Turing Machine, I can explain all of its capacities, but, lacking knowledge of its initial state, I may be unable to predict its behavior (Moore, 1956). Less interestingly, but just as important, some systems are simply intractable. We can explain the swirling trajectory of a falling leaf, but it would be hopeless to predict it.[2]

---

[1]I don't mean to suggest that DN theorists were confused about this. On the contrary, they held that explanation and prediction are just two sides of the same coin. The point is rather that DN conflates explanation and prediction, which are, I claim, orthogonal.

[2]Cartwright (1983) denies that we can explain the trajectory of a falling leaf. But all she argues for is that we cannot predict it. She seems to think it follows from this that we have no reason to believe that the laws of mechanics accurately subsume it.

Finally, many systems are well-understood in one or another idealized form, but their actual behavior cannot be predicted because the relevant boundary conditions are seldom or never realized.

So, systems can be well-understood yet unpredictable. What about the converse? Can a system be predictable without being understood? Certainly. For centuries, the tides have been predicted from tide tables (and continue to be). Their predictability was not improved at all by Newton's successful explanation of them.[3] Consider also the plight of the seventeenth century scientist confronted with the fact that pounding a nail makes it hot. Caloric theory, the going theory of heat at the time, treated changes in heat as diffusion phenomena. Your coffee cools because the caloric in it diffuses into the surrounding cup and air until equilibrium is reached. The fire reheats it because the caloric in the fire diffuses into the pot and surrounding air, and thence to the coffee, and so on. But pounding a nail will make it hot regardless of the temperature of the hammer.[4] This phenomenon -- call it the Galileo Effect

---

A more conservative view is that we understand falling leaves quite well. No one seriously thinks this is an outstanding mystery of nature on a par with the nature of consciousness, say. The problem is just that prediction is intractable.

[3]This is an interesting case in a number of ways. Newton's successful explanation in terms of the moon's gravitational influence does not allow prediction. This is done today, as before Newton, by tables. So here we have in a single instance a case in which prediction is neither necessary nor sufficient for explanation. Moreover, we have a case in which explanation seems to come apart from truth: The Newtonian mechanics on which the explanation is based have been supplanted, yet the explanation is still accepted.

[4]Friction was thought to release otherwise bound caloric, but this will not help with a cold hammer and nail.

after the man who made it famous -- is relatively easy to quantify. You can be in a position to predict what is going to happen, and even be able to quantify those predictions, yet still have no idea <u>why</u> it happens. Conversely, once in possession of the mechanical theory of heat, one sees that pounding a nail is like poking a cube of jello: more vibration equals more heat. But this insight doesn't improve predictability at all; it explains the Galileo Effect, but it is the statement of the Effect itself that generates the predictions.

3. <u>The laws of psychology are explananda because they specify effects</u>

From the perspective I've been urging, it emerges that a substantial proportion of research effort in experimental psychology isn't expended directly in the explanation business; it is expended in the business of discovering and confirming effects. An effect, I've been arguing, is an explanandum, not an explanans. In psychology, we are overwhelmed with things to explain, and somewhat underwhelmed by things to explain them with. Why is that?

I want to begin by mentioning a sociological factor just so it can be set to one side. The fact is that it is very difficult to publish a paper that simply offers an explanation of an effect. Most journals want reports of experiments. Explanation, such as it is, is relegated to the "discussion" section, which is generally lose and frankly speculative compared to the rest of the paper. Discussion sections are often not read, and their contents are almost never reported in other articles. The lion's share of the effort goes into the experiments and data analysis, not into explaining the effects they uncover. Any other course of action is a quick route to a plot in Tenure Memorial Park.

This is not mere tradition or perversity. It derives from a deep-routed uncertainty about what it would take to really explain a psychological effect. What, after all, would a successful explanatory theory of the mind look like?

We can be pretty sure what it wouldn't look like. It wouldn't look like a

<u>Principia Psychologica</u>. Newtonian Mechanics was laid out as an axiomatic system, self-consciously imitating Euclidian geometry, a widely influential paradigm in the Seventeenth Century, and has since been the dominant paradigm of an explanatory theory in science. It is arguable whether this is a really useful paradigm in any science. Certainly mechanics, even Newtonian mechanics, is never presented that way today. Still, if the goal is to lay out the fundamental principles of motion, the axiomatic approach makes a kind of sense. There are, one might suppose, a small number of fundamental principles governing motion, and these, together with some suitable definitions, might enable the derivations of equations specifying the (perhaps idealized) behavior of any particular mechanical system: a pendulum, a spring, a solar system, and so on. What makes this seem a viable approach is the idea that motion is the same everywhere, whatever moves, wherever and whenever it moves. It is also this sort of idea that grounds the widespread conviction that physics is the most fundamental science.

Conversely, what grounds the idea that psychology and geology are not fundamental sciences is the thought that psychological and geological systems are special. The principles of psychology and geology and the other so-called special sciences do not govern nature generally, but only special sorts of systems. Laws of psychology and geology are laws <u>in situ</u>, i.e., laws that hold of a special kind of system because of its peculiar constitution and organization. The special sciences do not yield general laws of nature, but rather laws governing the special sorts of systems that are their proper objects of study. Laws <u>in situ</u> specify effects--regular behavioral patterns characteristic of a specific kind of mechanism.

Once we see that the laws of a special science are specifications of effects, we see why theories in such sciences could not be anything like

Newton's <u>Principia</u>. Who would be interested in an axiomatic development of the effects exhibited by the liver or the internal combustion engine? What we want is an explanation of those effects in terms of the constitution and organization of the liver or engine. At the level of fundamental physics, laws are what you get because, at a <u>fundamental</u> level, all you can do is say how things are. We don't think of the fundamental laws of motion as effects, because we don't think of them as specifying the behavior of some specialized sort of system that behaves as it does because of its constitution and organization. The things that obey the fundamental laws of motion (everything) do not have some special constitution or organization that accounts for the fact that they obey those laws. The laws of motion just say what motion <u>is</u> in this possible world. Special sorts of systems, on the other hand, exhibit distinctive characteristic effects. In general, then, it seems that special sciences like psychology should seek to discover and specify the effects characteristic of the systems that constitute their proprietary domains, and to explain those effects in terms of the <u>structure</u> of those systems, i.e., in terms of their constituents (either physical or functional) and their mode of organization.[5]

4. <u>Effects and Capacities</u>

　　　What I have been calling psychological effects are not the only, or even the primary, explananda of psychology. I have been concentrating on effects because I have been criticizing the idea that psychological explanation is subsumption under law, and psychological laws specify effects. The primary explananda of psychology, however are not effects (psychological laws) but capacities: the capacity to see depth, to learn and speak a language, to plan, to

---

[5]See Cummins, 1983, chapters 1 and 2, for a detailed account of this kind of explanation in its application to psychology.

predict the future, to empathize, to fathom the mental states of others, to deceive oneself, to be self-aware, and so on. Understanding these sorts of capacities is what motivates psychological inquiry in the first place.

Capacities are best understood as a kind of complex dispositional property. Standard treatments typically assume that dispositions are specified by subjunctive conditionals along the following lines:

salt is water-soluble = If salt were put in water, then, cateris paribus, it would dissolve.

This sort of analysis is valuable because it makes it clear that to have a dispositional property is to satisfy a law in situ, a law characterizing the behavior of a certain kind of thing. Capacities and effects are thus close kin.

For this sort of analysis to work, we have to know what precipitating conditions (putting x in water) generate which manifestations (x dissolves). For many psychological capacities, it is a matter of some substance to specify exactly what they are. The specification of a capacity is what Marr (1982) called the "computational problem". This can be extremely non-trivial. How, after all, should we specify the capacity to understand Chinese? Or it can be relatively simple, as in the case of calculational capacities (the capacity to add or multiply, for example). So one reason we do not think of the capacity to learn a natural language as an effect is just that it is relatively ill-specified. As a consequence, the primary explananda of psychology--capacities--are not typically specified as laws.[6]

But there is a more interesting reason. Many of the things we call effects in psychology are in fact incidental to the exercise of some capacity of interest.

---

[6]It isn't clear that they always <u>can</u> be so specified. See the discussion of the capacity to play chess in section 6.2 below.

An analogy will help to clarify the distinction I have in mind. Consider two multipliers, M1 and M2. M1 uses the standard partial products algorithm we all learned in school. M2 uses successive addition. Both systems have the capacity to multiply: given two numerals, they return a numeral representing the product of the numbers represented by the inputs. But M2 also exhibits the "linearity effect": computation is, roughly, a linear function of the size of the multiplier. It takes twice as long to compute 24 X N as it does to compute 12 X N. M1 does not exhibit the linearity effect. Its complexity profile is, roughly, a step function of the number of digits in the multiplier.

The "linearity effect" is incidental to the capacity to multiply in M1. It is, as it were, a side effect of the way M1 exercises its capacity to multiply, and that is why we call this fact about computation time an effect and the multiplication a capacity. Of course, the "linearity effect" might be computed. We could design a system M3 that not only computes products, but computes reaction times as well, timing its outputs to mimic a successive addition machine. M3 might be quite difficult to distinguish from M1 on behavioral grounds, though it needn't be impossible. The timing function might be disabled somehow without disabling the multiplier. More subtly, computation of the relevant output times might itself be non-linear, in which case M3 will not be able to fool us on very large inputs (assuming it can process them at all). Or it might be that the "linearity effect" in M3 is cognitively penetrable (Pylyshyn, 1982), in which case it cannot be incidental. Thus, it can be a matter of substantive controversy whether we are looking at an exercise of a capacity or an incidental effect. This is precisely what is at issue between the friends of imagery and their opponents: are the rotation and scanning effects (for example) incidental effects of rotating/scanning a picture-like representation, or is it the exercise of a capacity to estimate rotation/scanning times involving

real physical objects? (See, for example, Pylyshyn, 1979.)

Thus, the primary explananda of psychological theory are capacities. Most capacities do not have to be discovered: everyone knows that people can see depth and learn language. But they do have to be specified, and that, to repeat, can be non-trivial. Effects are secondary explananda. They typically <u>do</u> have to be discovered. Much more importantly, however, is the different bearing explaining effects as opposed to capacities has on theory confirmation. Given two theories or models of the same capacity, associated incidental effects can be used to distinguish between them. This is important for two reasons. First, it is always possible in principle, and often in fact, to construct weakly equivalent models of the same capacity. To take an extreme case, Smolensky, Legendre and Miyata (1992) have shown that, for any parser written in a LISP-like language called TPPL (tensor produce programming language) it is possible to construct a distributed connectionist network that effects the same parses. With respect to parsing per se, then, there is nothing to choose between the two models. However, they predict very different incidental effects. Second, even when two models are not weakly equivalent, they may be on a par empirically: close enough so that differences between them are plausibly attributed to such factors as experimental error, idealization and the like. Again, incidental effects that may have no great interest as explananda in their own right may serve to distinguish such cases.

We can expect, then, to see a good deal of effort expended in the explanation of incidental effects that have little interest in their own right: no one would construct a theory just to explain <u>them</u>. But their successful explanation can often be crucial to the assessment of theories or models designed to explain the core capacities that are the primary targets of psychological inquiry.

5. Functional Analysis

  A theory may explain a dispositional property by systematic analysis--i.e., analyzing the system that has it, or it may proceed instead by analyzing the disposition itself. I call the application of property analysis to dispositions/capacities functional analysis.

  Functional analysis consists in analyzing a disposition into a number of less problematic dispositions such that programmed manifestation of these analyzing dispositions amounts to a manifestation of the analyzed disposition. By "programmed" here, I simply mean organized in a way that could be specified in a program or flow chart. Assembly line production provides a transparent illustration. Production is broken down into a number of distinct and relatively simple (unskilled) tasks. The line has the capacity to produce the product in virtue of the fact that the units on the line have the capacity to perform one or more of these tasks, and in virtue of the fact that when these tasks are performed in a certain organized way--according to a certain program--the finished product results. Schematic diagrams in electronics provide another familiar example. Since each symbol represents any physical object whatever having a certain capacity, a schematic diagram of a complex device constitutes an analysis of the electronic capacities of the device as a whole into the capacities of its components. Such an analysis allows us to explain how the device as a whole exercises the analyzed capacity, for it allows us to see exercises of the analyzed capacity as programmed exercises (i.e., organized) of the analyzing capacities.

  In these examples, analysis of the disposition goes together in a fairly obvious way with componential analysis of the disposed system, analyzing dispositions being capacities of system components. This sort of direct form-function correlation is fairly common in artifacts because it facilitates diagnosis

and repair of malfunctions. Form-function correlation is certainly absent in many cases, however, and it is therefore important to keep functional analysis and componential analysis conceptually distinct. Componential analysis of computers, and probably brains, will typically yield components with capacities that do not figure in the analysis of capacities of the whole system. A cook's capacity to bake a cake analyzes into other capacities of the "whole cook." Similarly, Turing machine capacities analyze into other turing machine capacities. Since we do this sort of analysis without reference to a realizing system, the analysis is evidently not an analysis of an realizing system but of the capacity itself. Thus functional analysis puts very indirect constraints on componential analysis. My capacity to multiply 27 times 32 analyzes into the capacity to multiply 2 times 7, to add 5 and 1, etc., but these capacities are not (so far as is known) capacities of my components.

The explanatory interest of functional analysis is roughly proportional to (i) the extent to which the analyzing capacities are less sophisticated than the analyzed capacities, (ii) the extent to which the analyzing capacities are different in kind from the analyzed capacities, and (iii) the relative sophistication of the program appealed to-i.e., the relative complexity of the organization of component parts/processes that is attributed to the system. (iii) is correlative with (i) and (ii): the greater the gap in sophistication and kind between analyzing capacities and the analyzed capacity, the more sophisticated the program must be to close the gap.

Ultimately, of course, a complete theory for a capacity must exhibit the details of the target capacity's realization in the system (or system type) that has it. Functional analysis of a capacity must eventually terminate in dispositions whose realizations are explicable via analysis of the target system. Failing this, we have no reason to suppose we have analyzed the capacity as it

is realized in that system.

6. <u>Existing explanatory paradigms in psychology</u>.

Here is the territory traversed so far:

1.   Psychological Explanation isn't subsumption under law.

2.   Psychological laws are not general laws of nature, but laws <u>in situ</u>, i.e., specifications of effects, not explanatory principles.

3.   The primary explananda of psychology are capacities.

4.   Effects and capacities in special kinds of systems are generally to be explained by appeal to the structure of those systems.

5.   Much of the effort in psychology, and almost all of the methodology, is devoted to the discovery and confirmation of effects.

It is striking that, while there is an extensive body of doctrine in psychology about the methodology appropriate to the discovery and confirmation of effects, there is next to nothing about how to formulate and test an explanation.[7] This is not surprising. If you think that explanation is subsumption under law, then you will see the discovery and testing of laws as the same thing as the formulation and testing of explanations. It may be a measure of the ubiquitousness of DN thinking that the methodology of hypothesis testing is nowhere complimented by a comparably sophisticated methodology of explanation testing. On the other hand, it may be that explanation testing simply doesn't admit of formulation in an explicit methodology because successful explanation has as much to do with the

---

[7]There is hypothetico-deductivism (HD): Explanations are <u>theories</u>, and theories of the sort in question are tested by deducing from them what effects should be exhibited. Explanations are then tested by determining whether the effects they predict are real.

knowledge and cognitive capacities of the explainers as it does with the logical properties of the explanation, a possibility I will return to below. Whatever the cause, psychologists faced with the task of explaining an effect generally have recourse to imitating one or another of the explanatory paradigms established in the discipline. These are familiar enough, but a brief review in the present context will prove illuminating.

There are five general explanatory paradigms that are influential in contemporary psychology. They are:

1. Belief-Desire-Intention (BDI) explanations;

2. Computational symbol processing explanations;

3. Connectionist explanations;

4. Neuroscience explanations;

5. Evolutionary explanations.

6.1 <u>Belief-Desire-Intention (BDI)</u>. This is by far the most familiar explanatory model. It is the model of common sense psychological explanation, as well as Freudian psychodynamics, and a great deal of current developmental, social and cognitive psychology. It is what Dennett praises as explanation from the intentional stance, and what Churchland deplores as folk-psychology. (Dennett, 1987; Churchland, 1981.) Underlying BDI is a set of defining assumptions about how beliefs, desires and intentions interact. These assumptions are seldom if ever made explicit. just as one does not make explicit the mechanical assumptions about springs, levers and gears that ground structural explanations of a mechanical machine. Everyone knows that beliefs are available as premises in inference, that desires specify goals, and that intentions are adopted plans for achieving goals, so it doesn't have to said explicitly (except by philosophers).

It is truly amazing how powerful this scheme of things is, particularly if

unconscious beliefs, desires and intentions are allowed. But there are problems. The most fundamental of these is something I call Leibniz' Gap. Here is Leibniz' formulation of the Gap.

> Moreover, we must confess that the perception, and what depends on it, is inexplicable in terms of mechanical reasons, that is, through shapes and motions. If we imagine that there is a machine whose structure makes it think, sense, and have perceptions, we could conceive it enlarged, keeping the same proportions, so that we could enter into it, as one enters into a mill. Assuming that, when inspecting its interior, we will only find parts that push one another, and we will never find anything to explain a perception. And so, we should seek perception in the simple substance and not in the composite or in the machine. (Leibniz, sec. 17)

There is, as Liebniz points out in this famous passage, a gap between the concepts of BDI psychology, and those we use to describe the brain. So, even if we are convinced that the mind is the brain, or a process going on in the brain, physical observation of the brain seems to give us data in the wrong vocabulary: synapses rather than thoughts. When we look at a brain, even a living brain, we don't see thoughts. Or, not to beg the question, we don't see anything we readily recognize as thoughts. If you had a Newton camera and took a snap shot of a billiard game in progress, you would see vectors with centers of gravity at their tails. If you had a psychology camera and took a snap shot of a living brain, you would, according to BDI psychology, see beliefs, desires, intentions and their canonical relations. But to build a psychology camera, you would need to somehow bridge Liebniz' Gap by correlating observed brain properties, events and processes with beliefs, desires and intentions, and this, at least for now, is beyond us. Thus, the wide Leibnizian gap between BDI psychology and the brain is destructive to satisfying

psychological explanation. Lacking some precise suggestion about how beliefs, desires and intentions are instantiated in the brain, we are left wondering whether even the most powerful BDI analysis of some psychological effect might specify a way to achieve the effect, but not the way, i.e., the way the brain does it. This objection is a "philosophical" objection in that it is independent of how predictively successful BDI analyses turn out to be. If we knew there was only one way to achieve the psychological effects we find in the laboratory and in the field, then the fact that a psychological effect had a satisfactory BDI analysis would constitute evidence that the brain must somehow realize the structure that analysis specified. But, of course, we don't know that there is only one way to design a mind like ours, and, lacking this knowledge, we don't know whether the predictive inaccuracies that accompany any scientific theory are due to the fact that the human mind is not a BDI device or to the fact that our theory is idealized, that measurement isn't perfect, and so on.

Another serious conceptual problem with BDI has to do with the nature of the propositional attitudes --belief, desire, intention and their kin -- that are the workhorses of BDI. BDI psychology requires a conception of the attitudes that allows for beliefs, desires and intentions that are not only not conscious, but which cannot be made conscious. While most philosophers and psychologists find this acceptable, it has not gone unchallenged (Searle, 1992). Somewhat more seriously, BDI requires that the attitudes be "atomistic", i.e., that they can exist in relative isolation. In a BDI framework, standard accounts of linguistic or visual processing, for example, require beliefs about phrase structures and zero-crossings in subsystems that are relatively isolated informationally from other aspects of cognition. No psychologist working on concepts and their acquisition would think that merely being able to see, or to understand language, is

sufficient for having the concept of a phrase structure or a zero-crossing. Yet having beliefs about phrase structures and zero-crossings seems to require having these concepts. Thus, atomism about the attitudes, though it has its defenders (Fodor and Lepore, 1992) is by no means uncontroversial (Stich, 1983; Block, 1986).[8]

Finally, it isn't clear that psychological phenomena generally can be reduced to the interaction of propositional attitudes, even if these are broadly construed to include such things as the language processor generating a representation of the phrase structure of the current linguistic input. BDI seems best suited to so-called "higher" cognition, and in particular, to high-level reasoning and planning. Even here, there are formidable critics. Eliminativists have argued that BDI, whether it be "folk theory" or grounded in an innate theory of mind, is, in fact, discredited theory (Churchland, 1981).

6.2 <u>Computationalism</u>. Computationalism (the brain is a computer and the mind is what it is doing) is just BDI minus some of the baggage. Computationalism is a "top down" strategy. In the hands of the computationalist, that strategy begins by identifying a task or capacity to be explained: the capacity to learn a language, or converse, or solve a problem, etc. It then attempts to specify that capacity as a function or relation: What inputs produce what outputs under what circumstances? Finally, that characteristic function or relation is analyzed into components that have known

---

[8]It is interesting that, as the phenomena become more "specialized", intention and desire tend to drop out. There is surely some truth in the idea that the game of life is to form intentions (plans) that will get things moved from the desire box (Desire[I am rich]) to the belief box (Believe[I am rich]). But it is a stretch to think that this is the fundamental loop in language processing or vision.

computational realizations. (In practice, this means analysis into components that can be programmed in LISP or some other standard programming language.)

This strategy involves three assumptions and a precondition that are worth noting.

1. One underlying assumption of this approach is that psychological functions are computable. This is actually a rather strong and daring assumption. Most dynamical systems found in nature cannot be characterized by equations that specify a computable function. Even three bodies moving in Newtonian space do not satisfy this assumption. It is very much an open question whether the processes in the brain that subserve cognition can be characterized as the computation of a computable function.

2. Another underlying assumption of top-down computationalism as it is usually characterized (and as I have just characterized it) is that psychological capacities <u>can</u> be specified independently of their analyses. But this is pretty patently false in many cases: There is, for example, no input-output function the computation of which would constitute playing intelligent chess. Or rather, there are a great many. Think of a chess system as a move generator, i.e., as a function from board positions (current) to board positions (the move). In a given situation, intelligent chess players might make any number of different moves. Indeed, the same one might make different moves on different occasions. In practice, then, the only way to specify a chess function is to actually write an algorithm for computing it. We cannot, in general, expect to specify a cognitive function before we analyze and implement it, and this introduces a methodological difficulty: If we cannot specify the explanandum independently of the explanans, how are we to compare competing explanations? We can, of course, determine which theory better predicts

whatever observational data there is--i.e., we can determine which does a better job predicting whatever known effects there are--but this tells us only which underlying theory is more likely true, not which generates the better explanation. The distinction is important. It is well known that if it is possible to accommodate the data at all, it is possible to accommodate it with a theory that says nothing whatever about the underlying mechanisms or their analysis, i.e., in a way that has no explanatory force whatever (Craig, 1953; Putnam, 1965). This problem is underappreciated because a tendency to focus exclusively on accommodating effects leaves explanatory issues out of the picture from the start.

3. A third underlying assumption of the top-down strategy, closely related to the second assumption, is that we will be able to recognize and characterize the relevant inputs and behaviors antecedently to serious attempts to explain how the later are computed from the former. Here the difficulty is that pre-analytic conceptions of behavior and its causes may seriously misrepresent or distort what is actually going on. Connectionists sometimes complain that there is no reason to think that cognition in the brain is the manipulation of representations that correspond to our ordinary concepts. Top-down strategists therefore run the risk of characterizing the explananda in terms that are cross-cut or distort the categories that are actually causally relevant. This is analogous to the almost irresistible temptation in biology to believe that the morphological traits of importance and interest to us must correspond to our genes in some neat way. Computationalists are wont to reply that what Daniel Dennett (1987) calls the intentional stance--predicting and explaining behavior in terms of beliefs, desires and intentions--is enormously successful, and hence that it cannot be fundamentally wrong to characterize cognition in something like these common sense terms. The same can be said

for Ptolemaic astronomy, or Newtonian mechanics, however. Considerable explanatory and predictive success is possible with fundamentally mistaken or even incoherent theory.

So much for the assumptions. Now for the precondition: A successful application of the top-down strategy requires that the target explanandum is analyzable. Everyone who has ever tried their hand at programming is familiar with this constraint. You cannot write a program that computes bids in bridge, or computes square roots, if you do not know how to compute bids in bridge or compute square roots. But many psychological capacities are interesting explananda precisely because we have no idea how the task is done. This is why artificial intelligence plays such a central role in computationalism. It requires very considerable ingenuity to discover a way--any way--to construct 3D specifications of visual space from retinal images, or to make it happen that, in problem solving, two short sessions are more effective than one long one.

But even with success, there is a problem: having figured out <u>a</u> way to compute a cognitive function, what reason is there to think that that is how our brains do the job? I do not mean to suggest that there is no way of addressing this problem, only that it is a problem that is bound to arise in a top-down framework. Computationalists are thus inevitably left with a narrowed but still substantial Leibnizian Gap: the gap between a computational description of psychological processes and a bio-neural description of the processes in the brain.[9]

---

[9]The gap is narrowed relative to BDI because a computational analysis will at least have demonstrated the physical--indeed computational--realizability of the processes they postulate. BDI explanations are always subject to the eliminativist worry that the fundamental processes postulated have no physical realizations at all. Still, it is arguable that many computationalist explanations only make sense on the controversial assumption that beliefs, desires and

6.3 <u>Connectionism</u>. The top down strategy is explanandum driven: you begin with a capacity to explain, and try to find a computational architecture that will have it. The bottom up strategy is explanans driven: you start with a specification of the architecture, and try to find a way to make it do the task.[10] What connectionists have in common is the assumption that cognitive capacities are built out of a stock of primitive process designed explicitly to be rather brain-like. They begin with the building blocks of a simplified and idealized brain, and attempt to create systems that will behave in a recognizably cognitive way. The connectionist thus seeks to narrow the Leibnizian Gap even further to that between a genuinely bio-neural description of the brain, and the simplified and idealized "neural networks" that are their stock in trade.

But a much narrowed Gap is not the only payoff. As it happens, it is possible to program connectionist networks to do tasks that the programmer does not know how to do. All that is required is a sufficiently representative "training set": a set of inputs paired with their correct responses. Thus the precondition of top-down computationalism discussed above can be avoided. You can program a network to do a task you haven't the faintest idea how to do. There is a downside to this, however: once you have trained a network, you may still have little if any idea how it does the task. Studying an artificial

_____

intentions have reasonably straightforward computational realizations. I return to this point below.

[10]In practice, most computationalist are actually bottom-uppers to some extent. This is because, as a graduate student, you apprentice in a research group that is more or less committed to a given architecture, and you job is to extend this approach to some new capacity. It is just as well: pure top-downism, as described by Marr (1982), is probably impossible. Computationalist architectures, however, are not well-grounded in the brain, so the problem just rehearsed remains.

network is, of course, much easier than studying a living brain, so you are still substantially ahead. But you are not home free.

Moreover, it is seldom noticed that one of the lately discussed assumptions required by the top down approach is also required by bottom-uppers. Training sets must be specified somehow, and the problem of how to conceptualize inputs and behaviors is no easier for connectionists than it is for top-down computationalists. While connectionists need not assume that networks operate on internal representations that correspond to ordinary common-sense concepts, they are no better off than top-down computationalists when it comes to conceptualizing the target explananda.

Before we leave the topic of underlying assumptions and enabling conditions, it is worth pausing to note that some of the central enabling assumptions of computatinalism are shared by connectionism. Both assume the possibility of Spock--i.e., that the mind is basically a cognitive engine and only secondarily a seat of emotion, feeling and sensation. Both assume that consciousness is inessential to the understanding of cognition. And both assume that cognition doesn't require a biological brain, let alone an immaterial soul. Both are thoroughly functionalist and materialist. And both are representationalist in that both assume that cognition is to be understood as disciplined transformation over states whose primary function is the representation of information relevant to the cognitive capacity being exercised. The differences that divide computationalist and connectionist are practically invisible against the scale that measures the distance between them and the behaviorism of Watson or Skinner, or the structuralism of Titchner.

6.4 <u>Neuroscience</u>. Everyone who is not a dualist believes that mental processes are process that go on in the brain. If one's goal is a science of the mind, however, observation of the brain seems to yield results on the wrong

side of Leibniz' Gap. The Computationalist response to this problem is to try to understand cognitive processes in abstraction from the brain or any other "hardware" in which they might occur. The Computationalist strategy is to first articulate a computational theory of cognition, and then to inquire into how the implicated computational processes might be carried out in the brain. This strategy has some evident merits. Since no one doubts that computational processes can be physically realized, Computationalism is free from any dualist taint. Yet the problem of bridging Leibniz' Gap is conveniently put off until some future date when we will surely know more about both cognitive and neural processes. An evident drawback, however, is that there is no guarantee that cognitive processes are computational processes at all, let alone that cognition in biological brains will turn out to be the kind of processes we are led to investigate by following a strictly top-down approach. Although that approach has had some notable successes, it has also had some notable failures. It would not be unreasonable to conclude that the difficulties faced by Computationalism might be due to insufficient attention being paid to the only processes we know for sure are sufficient to subserve mentality in general, and cognition in particular, namely brain processes. Perhaps we should simply accept the fact that, as things currently stand, studying the brain puts us on the wrong side of Liebniz' Gap, but hope that, as our knowledge increases, the outlines of a bridge over the Gap will eventually appear.

Connectionists attempt to take a middle ground here, starting in the middle of the Gap, as it were, and trying simultaneously to bridge to either side. Most neuroscientists, it seems, are at least tolerant of the Connectionist strategy. But they are inclined to argue that connectionist models are such vastly oversimplified models of the brain as to be very likely misleading at best. If we are going to bridge Liebniz' Gap, we are going to have to know a great

deal more about the brain than we do now. This much is agreed on all hands. So why not get on with it? And, since the brain is the only known organ of mentality, whether natural or artificial, it seems only sensible to begin by trying to understand how it works. Any other strategy arguably runs the risk of being a wild goose chase, an attempt to make mentality out of stuff that just isn't up to the job.

This line of argumentation has been around at least since the seventeenth century, but it had few practical consequences until relatively recently simply because there was no very good way to study the brain. Steady technological progress, however, is beginning to make Leibniz' thought experiment a reality. As a result, the problem he articulated so eloquently is forced upon us anew, for, marvelous as the new technology is, it does not, and cannot, provide "psychology glasses", lenses through which observed brain anatomy and activity emerge as psychological faculties and thought processes.

Technology can take us to the brink of Leibniz' Gap, but only theory can bridge it. There are two conceptions of how neuroscience might contribute to the bridge. According one approach, concepts generated by neuroscience proper to articulate its data and theory should be use to reconceive the mental from the bottom up, discarding those mentalistic concepts that have no clear neuroscientific reconstruction, and simply replacing those that do. (Churchland 1987). Psychology on the mental side of Leibniz' Gap will either be assimilated or perish. Well-confirmed effects remain as explananda on this view modulo the caveat that the concepts used in their articulation must not be tainted to deeply by concepts that have no acceptable neuroscientific reconstruction.[11]

---

[11]The history of science is full of effects that weren't real in the sense that subsequent science rediagnosed the inevitable failures to fit the data precisely as conceptual error rather than experimental error.

Psychological capacities of the sort that constitute the primary explananda of more top down approaches are viewed with suspicion; guilty (until proven innocent) of not cutting nature at the joints. I'll call this the strong neuroscience program.¹²

As things stand, the strong neuroscience program is almost impossible to put into practice. Standard descriptions of dissociations, of tasks done during FMRI, and so on are up to their eyebrows in terminology from the "wrong" side of Leibniz' Gap. A more common and more ecumenical conception of the role of neuroscience treats it as a source of evidence designed primarily to arbitrate among functional analyses formulated in other terms, terms from unreduced psychology residing on the other side of the Gap from "pure" neuroscience. There are serious methodological issues here that are matters of controversy in psychology, neuroscience and philosophy, but it is clear in a general way how weak neuroscience bears on the issue of psychological explanation: it passes the buck. On this conception, psychological effects and capacities are explained as the effects or capacities o f BDI, computationalist or connectionist systems, and these are assumed to be instantiated somehow in the brain. Neuroscience enters this picture as a source of evidence arbitrating among competitors, and ultimately, as the source of an account of the biological realization of

---

¹²The use of "strong" and "weak" to distinguish two conceptions of the role of neuroscience in psychological explanation, and the use of these words to distinguish two analogous conceptions of the role of evolutionary theory in psychological explanation, should not be taken as terms of approbation or abuse. They are modeled after Searle's well-known distinction between strong and weak AI (Searle, 1980).
    Perhaps I should emphasize as well that I am not here attempting to characterize neuroscience, but only its abstract role in psychological explanation. The same goes for my remarks about evolution in the next section.

psychological systems described functionally.

6.5 <u>Evolutionary Explanations</u>. Like neuroscience, evolution can be regarded as either a source of psychological explanations or as a source of evidence bearing on one or another non-evolutionary theory that generates its own psychological explanations, and this generates a distinction between a strong evolutionary program and a weak evolutionary program analogous to the distinction between the strong and weak neuroscience programs. The evidential role of evolution is relatively easy to specify. Functional analyses attribute functions to the analyzed systems. A source of evidence that a system really has a given function, or has a component with a given function, is that such a function would have constituted an adaptation, or the likely corollary of an adaptation, for the system's ancestors.[13] Conversely, a functional analysis that proposes functions in a biological system that have no plausible evolutionary rationale are suspect on the grounds that nature is not being carved a the joints. Again, there are important methodological issues here, but they don't bear on the nature of psychological explanation, only on the confirmation of the theories that generate them.

The strong evolutionary program is based on the idea that evolution might actually explain a psychological capacity or effect. This idea is difficult to articulate and assess. At best, it seems that evolution might explain why a certain psychological capacity or effect is pervasive in a given population. It

---

[13]Corollary: x was an adaptation, and y is a likely precondition or consequence of having x, so whatever evolutionary argument exists for x confers some plausibility on y as well.

I don't mean to suggest that adaptation and selection is all there is to evolution. But non-selectionist scenarios for the evolution of a psychological function are bound to be relatively difficult to construct or confirm.

could, to put it crudely, explain <u>why</u> we see depth, but not <u>how</u>. Thus, an evolutionary explanation and an explanation generated by one of the other paradigms would not be direct competitors in the same explanatory game. This is obscured by the fact that evolutionary reasoning could favor some functional analyses over others, a fact that entails that evolutionary explanations could be incompatible with explanations generated by one of the other frameworks (BDI, Computationalism, Connectionism, Neuroscience). But evolutionary explanations do not seek to answer the same question as those generated by the other frameworks. Hence, as long as there is no incompatibility in the functional analyses each postulates, there is no reason why we should have to choose between an evolutionary explanation and, say, a connectionist explanation or a BDI explanation.

 7. <u>Two Problems for Psychological Explanation</u>

      The first three of the familiar frameworks just rehearsed--BDI, Computationalism, and Connectionism-- are, as they should be, <u>analytical</u> frameworks. That is, they are frameworks for analyzing (decomposing) complex capacities into more primitive components. The strong neuroscience program aspires to be an analytical framework, and is, perhaps, well on the way to becoming one. Weak neuroscience and the weak evolutionary program, do not pretend to be explanatory frameworks in their own right, hence offer no alternative to the analytical approach. Finally, what I have called the strong evolutionary program is, I think, best construed at explaining the prevalence of an effect or capacity in a population, and thus leaves untouched the question of what the mind is and how it works.

      Our survey of the currently viable explanatory frameworks thus reveals that, although there is still considerable lip service paid to DN, actual theory building and explanation construction takes place in frameworks that are not

designed for the elaboration of laws but rather are designed for the elaboration of functional analyses. The foundational problems for psychological explanation, then, are special versions of the problems that arise for functional analysis generally. If we leave aside strictly epistemological problems, problems about how functional analyses are to be "discovered" or confirmed, and focus solely on how they work as explanations, two central issues emerge.[14] The first might be called the <u>realization problem</u>. Functional analysis always leaves one with a gap between the functional characterization of a system and the various non-functional characterizations that are assumed to apply to the system whose functional analysis is at issue.[15] In psychology, this is what I have called Leibniz' Gap. The second problem might be called the <u>unification</u> <u>problem</u>. What I have in mind here is the fact that functional analyses are usually generated to explain some particular capacity or effect, or a closely related set of them. Researchers concerned with some aspect of vision may be sensitive to the issue of unifying their account with those directed at some other aspect of vision. But they are

---

[14]I don't mean to suggest that these problems are trivial or unimportant. Indeed, I think they are many and deep. But these are problems about confirmation, not about explanation. One of the many unfortunate consequences of DN is that it (intentionally!) blurs the distinction between confirmation and explanation.

[15]As many have pointed out (see, for example, Lycan, 1987), the distinction between functional and non-functional levels of organization is relative. Realizing systems are seldom characterized in non-functional terms. They are rather characterized in terms of functions that differ from those whose realization is at issue. A logic circuit, for example, might be analyzed in terms of AND gates, OR gates, and INVERTERS. The realization of this circuit might then be specified in terms of resistors and transistors and conductors. These are themselves, of course, functional terms, but their realization is not at issue, so they count as non-functional relative to the gates and invertors whose realization is being specified.

less likely to concern themselves with making their analyses fit with those being generated by those researching language or emotion or reasoning.

7.1 <u>Leibniz' Gap: Intentionality and Consciousness</u>. The realization problem, in the form of Leibniz' Gap, looms for every current explanatory framework surveyed above with the exception of strong neuroscience which holds that concepts not proprietary to neuroscience itself need not be taken seriously. While attractive philosophically because it eliminates the Gap, strong neuroscience is, as remarked above, nearly impossible to put into practice as an explanatory strategy simply because the vast majority of the explananda are formulated in terms that either explicitly or implicitly draw on concepts that have no known counterparts in neuroscience. Neuroscience that does honor eliminativist constraints seems, at present anyway, to have little to do with psychology. I propose, therefore, to put the strong neuroscience program aside and concentrate on those frameworks that must, in one way or another, face Leibniz' Gap.

There is no special mystery about what counts as a satisfactory solution to realization problems generally. Every time we design an artifact to satisfy a functional characterization and then build it we solve a realization problem. This shows that there is no special philosophical mystery about what it is to realize a functionally specified system. Difficulties arise, however, in special cases in which there is a fundamental unclarity in one or more of the primitives of the analytical framework. There is deep uncertainty about whether beliefs, desires and intentions can be computationally realized, not because we do not understand what realization requires, but because we are unclear about beliefs, desires and intentions. There is no comparable worry about whether a given computationally specified system is realized in the brain. There is uncertainty, of course, but it is a different kind of uncertainty. We know what it takes to realize

a computationally specified system, we just don't know if what it takes is in the brain. But we don't know what it takes to realize a belief or desire.[16] Do any of the many sophisticated planners currently in the literature actually have beliefs, desires and intentions? And if they do not, should we conclude that planning doesn't require belief, desire and intention, or should we conclude that computationalist planners are mere imitators of mental activity? Everyone recognizes these as Philosophical Questions, which, in this context anyway, means mainly that everyone recognizes that they are questions that, as things now stand, cannot be addressed experimentally. And, of course, there is an exactly parallel, and perhaps related (Searle, 1992), set of problems about consciousness.

It is important to see that the Liebnizian Gap between intentional states like belief, desire and intention on the one hand, and computationalist, connectionist or neuroscience concepts on the other is not just a problem for BDI. It is a problem for any framework that either characterizes its explananda in intentional terms, or assumes (tacitly or explicitly) a realization of intentional states in its proprietary mechanisms and processes whether these be the computational manipulation of data structures, the spread of activation disciplined by connection weights, or synaptic connections and spiking frequencies. I think it is pretty obvious that both kinds of intentionalist taint are ubiquitous, though not universal, in psychology and artificial intelligence. I submit that this is why so much psychological explanation, while it is often compelling and informative, is almost always ultimately unsatisfying. What is more, we don't know whether the problem is just that we don't really understand intentional states, or that, as eliminativists claim, there is nothing to

---

[16]Except trivially: a normal brain. All this does is rule out dualism.

be understood. We never solved the realization problem for entelichies either, but that was a knock on vitalism, not a failure of philosophical analysis.

All of this is old news, of course. But it is worth reminding ourselves that there is nothing wrong with psychological explanation that a solution (or dissolution) of the problem of intentionality and consciousness wouldn't cure.

7.2 <u>The Unification Problem</u>. Well, almost nothing. There is a <u>de facto problem</u> that plagues psychological explanation, and that is its evident lack of unification.

The first and most obvious problem is that there are four quite different explanatory frameworks operative in contemporary psychology: BDI, Computationalism, Connectionism, and (strong) Neuroscience. While the first two and the second two are reasonably close together, it remains true that explanations constructed in one framework are seldom translatable into explanations in another. The gap between BDI and Computationalism on the one hand, and Connectionism and (strong) Neuroscience on the other is particularly wide and typically competitive.

It is a commonplace in science to attack different problems from the perspective of different explanatory models. To explain the flow of water and wave propagation, one typically models water as a continuous incompressible medium. To explain diffusion and evaporation, one models water as a collection of discrete particles.[17] But it is important to see how this situation differs from the situation that prevails in psychology. The different models of water are rung in to explain different effects. While water cannot be both a continuous incompressible fluid and a cloud of free molecules, each model is directed at a different set of problems. There is no competition between the models

---

[17]The example is from Paul Teller, in conversation.

concerning the solution of the <u>same</u> problem.[18] In contrast, it is notorious that connectionist and computationalist models compete in just this way, a classic example being the explanation of the acquisition of the past tense in English (Rumelhart and McClelland, 1986; Pinker and Prince, 1989). In this respect, contemporary psychology resembles 17th C. mechanics in which Cartesians and Newtonians competed to explain the same phenomena within different frameworks. There is, of course, no way to resolve this kind of competition other than to let the science take its course. In the meantime, however, <u>every</u> explanation is, to some extent, undermined by the deep disunity that afflicts the field in its current state of development. Until the field is unified in some way--by the victory of one of the current competitors, by the emergence of a new framework, or by a successful realization hierarchy (BDI realized computationally realized as a connectionist network realized in the brain)--the suspicion remains that some or all of the explanations currently on offer are fundamentally flawed because they are articulated in a fundamentally flawed framework.

In addition to the disunity across frameworks, there is considerable disunity within each framework, particularly within computationalism and connectionism.[19] Both frameworks allow for an enormous variety of models

---

[18]I don't mean to suggest that this situation is entirely unproblematic. It is certainly tempting to suppose that there is a deep disunity here unless both models of water can be treated as acceptable idealizations or simplifications grounded in a deeper single model.

[19]Functional analyses tend to proliferate when there are no strong restrictions on the primitives. Computationalism, in principle, allows any computable function as a psychological primitive. Connectionism is somewhat less permissive, but there is still a bewildering variety of network architectures currently on offer. Strong Neuroscience, in so far as it exists as an explanatory framework at all, enforces very little in the way of architectural constraints

based on very different principles.[20] Attempts at unity are not unknown: In the computationalist camp, Anderson's ACT* and Newell's SOAR spring to mind, as does Grossberg's ART in the connectionist camp. (Grossberg, 1982; Newell, 1990; Anderson, 1996). But it is an understatement to say that these are not widely accepted. The prevailing bewildering diversity of models tends to undermine confidence in any.

Having said all of this, I do not think we should worry much about disunity. The ordinary practice of good science will take care of disunity eventually. There is a far greater danger in forcing more unity than the data warrants. Good experimentation, like good decision making generally, can tell us which of two models is better, but it cannot tell us how good any particular model is. The best strategy, then, is to have a lot of models on offer on the grounds that, other things equal, the best of a large set is likely better than the best of a small one.

8. Conclusions

I have been urging that explanation in psychology, like scientific explanation generally, is not subsumption under law. Such laws as there are in psychology are specifications of effects. As such, they do not explain anything but themselves require explanation. Moreover, though important, the

beyond those dictated by gross anatomy.

BDI is probably the most unified of the currently viable explanatory frameworks, since it is defined by a choice of primitives. Still, there have been few systematic attempts to make the principles of interaction among these principles explicit. An exception is Freudian psychodynamics: but while this is (extended) BDI, but most of its fundamental principles--e.g., of repression-- would be regarded as dubious by many BDI researchers.

[20]As Smolensky, Legendre and Miyata (1992) have pointed out, explanation in these frameworks tends to be model based rather than principle based.

phenomena we typically call effects are incidental to the primary explananda of psychology, viz., capacities. Capacities, unlike their associated incidental effects, seldom require discovery, though their precise specification can be non-trivial. The search for laws in psychology is therefore the search for explananda, for it is either the search for an adequate specification of a capacity or for some capacity's associated incidental effects. Laws tell us what the mind does, not how it does it. We want to know how the mind works, not just what it does.

Capacities and their associated incidental effects are to be explained by appeal to a combination of functional analysis and realization, and the currently influential explanatory frameworks in psychology are all frameworks for generating this sort of explanation. Thus, in spite of a good deal of lip service to the idea that explanation is subsumption under law, psychology, though pretty seriously disunified, is squarely on the right track. Its efforts at satisfying explanation, however, are still bedeviled by the old problems of intentionality and consciousness. This is where psychology and philosophy meet. Psychology needn't wait on philosophy, however. The life sciences made a lot of progress before anyone knew how life is realized.

References

Anderson, J. 1996. <u>The Architecture Of Cognition</u>. Mahwah, NJ: Lawrence
    Erlbaum Associates.

Block, N. 1986. "Advertisement for a Semantics for Psychology." *Studies in the Philosophy of Mind*. In P. French, T. Uehling, Jr, and H. Wettstein, eds. Vol. 10 of *Midwest Studies in Philosophy*. University of Minnesota Press.

Cartwright, Nancy 1983. How the laws of physics lie. Oxford : Clarendon   Press ; New York : Oxford University Press.

Churchland, P. M. 1981. "Eliminative materialism and propositional   attitudes." The Journal of Philosophy, 78:67-90.

Craig, W. 1953. "On axiomatizability within a system." Journal of Symbolic   Logic, 18(1):30-2.

Cummins, R. 1983. The Nature of Psychological Explanation. Cambridge,   MA: MIT Press (a Bradford Book).

Dennett, D. 1987. The Intentional Stance. Cambridge, MA: MIT Press (a   Bradford Book).

Fodor, J. and Lepore, E. 1992. *Holism: A Shopper's Guide.* Oxford: Basil   Blackwell.

Garcia, J. and Koelling, R. 1966. "The relation of cue to consequence in   avoidance learning. Psychonomic Science 4:123-4.

Grossberg, S. 1982. Studies Of Mind And Brain : Neural Principles of   Learning, Perception, Development, Cognition, and Motor Control.   Dordrecht, Holland: D. Reidel.

Hempel, C. and Oppenheim, P. 1948. "Studies in the Logic of Explanation."

    Philosophy of Science, 15:135-75.

Hempel, C. 1966. Philosophy of Natural Science. Englewood Cliffs,
        N.J.:Prentice Hall.

Kim, J. 1962. "On the logical conditions of deductive explanation." Philosophy
of Science, 30:286-91.

Leibniz, G. (1714). The Monadology. Translation by R. Ariew and D. Garber
        (1989) Leibniz: Basic Works. Indianapolis, IN: Hackett.

Lycan, W. 1987. Consciousness. Cambridge, Mass.: MIT Press (a Bradford
        Book).

MacDonald, J. and McGurk, H. 1978. "Visual influences on speech perception
        processes." Perception and Psychophysics. 24 (3): 253-257.

Madigan, S. 1969 "Intraserial repetition and coding processes in free
        recall." Journal of Verbal Learning and Verbal Behavior, 8:828-835.

Marr, D. 1982. Vision. New York: Freeman

Moore, E. 1956. "Gedanken experiments on sequential machines." In C. Shannon
and J. McCarthy, Automata Studies, Princeton: Princeton University Press. pp.
129-156.

Newell, A. 1990. Unified Theories Of Cognition Cambridge, MA: Harvard

University Press.

Pinker, S., & Prince, A. (1989) "Rules and Connections in Human Language."
    In R. Morris (ed.), Parallel Distributed Processing (pp. 182-199). Oxford:
Oxford University Press.

Putnam, H. 1965. "Craig's theorem." The Journal of Philosophy, 62(10).

Pylyshyn, Z. 1979. "The rate of "mental rotation" of images: A test of a holistic
analogue hypothesis." Memory and Cognition, 7(1):19-28.

Pylyshyn, Z. (1982). Computation and Cognition. Cambridge, MA: MIT Press
    (a Bradford Book).

Rumelhart, D., & McClelland, J. 1986. "On Learning the Past Tenses of
    English Verbs." In J. McClelland, D.E Rumelhart, and the PDP Research
Group (eds.), Parallel distributed processing, volume 2 (pp. 216-271).

Salmon, W. 1998. Causality and Explanation. New York: Oxford University Press.

Searle, J. 1980. "Minds, Brains, and Programs." Behavioral and Brain
     Sciences. 3: 417-24.

Searle, J. 1992. The Rediscovery of Mind. Cambridge, MA: MIT Press.

Smolensky, P., LeGendre, G., and Miyata, Y. 1992. "Principles for an integrated
connectionist/symbolic theory of higher cognition." Tech. Report 92-08,

Institute of Cognitive Science, University of Colorado.

Stich, S. 1983. *From Folk Psychology to Cognitive Science: The Case Against Belief.* Cambridge, MA: MIT Press (a Bradford Book).