Functions and Mechanisms in Contemporary Neuroscience[1]

Carl F. Craver
PNP Program
Washington University, St. Louis

§ 1 Introduction. Functional descriptions occupy a central place in the theories of contemporary neuroscience. These theories contain such conspicuously functional nouns as "vesicle," "neurotransmitter," "receptor," "channel" and "ocular dominance column." And the entities that they designate are said to "dock," "signal," "recognize," "modulate," and "filter." Many of the adjectives are functional too: there are *inhibitory* neurotransmitters, *voltage sensitive* channels, *regulatory* peptides and *feedback* loops. As for explicit commitment to functional description, one need look no further than the central, if controversial, dogma of contemporary neuroscience that *functions* localize to more or less well-circumscribed regions of the brain. What do we assert of an item when we assign it a function? What sorts of evidence can be used to justify functional attributions? And what role do functional attributions play within the predominantly mechanistic explanatory framework of the contemporary neurosciences? Here I suggest a unified answer to these questions.[2]

Philosophers of biology have puzzled over the legitimacy of functional description in a mechanistic world (e.g. Rosenberg, 1985) and over the ability of functional description to support normative claims: claims about how some item ought to

[1] Thanks to Lindley Darden, Peter Machamer, Greg Morgan, Bob Olby and Ken Schaffner for substantive comments concerning the structure and content of this essay. An early draft of this paper was produced as the Cognitive Studies Post-Doctoral Fellow at the University of Maryland.
[2] For similar attention to the role of functional statements in more physiologically inclined biological sciences, see Wouters (1999).

behave or about the item's goal. But the notion of "function" as it appears in contemporary neuroscience is seamlessly integrated into the mechanistic structure of its theories. Furthermore the kinds of functional attributions that litter the theories of neuroscience are neither intended to distinguish between how an item ought to behave from the way that it in fact behaves nor to uncover the goal towards which the item strives. Instead, functional description serves the explanatory role of showing how an item fits into a multilevel mechanism. The theories of contemporary neuroscience describe mechanisms, and functional attribution is a means of integrating item into a hierarchical nexus of mechanisms. Such integration is explanatory (in Salmon's 1984 sense of causal-mechanical explanation), and it is also useful in the process of mechanistic theory building. Functional attributions are contentful to the extent that they can be cashed out in a detailed description of how an item is organized into a mechanism, and one has good evidence for one's functional description when one can show that the item is organized into the mechanism in precisely that way. In what follows, I exhibit and elaborate these claims with the help of some examples of functional attribution in the recent history of neuroscience.

2 Explanation, Mechanisms and Levels. Explanations in the neurosciences typically come in the form of descriptions of mechanisms. Wesley Salmon elegantly characterized causal mechanical explanation as a matter of showing *how* some phenomenon or type of phenomenon fits into the *causal* nexus. He contrasted his statement of causal mechanical explanation with Hempel's Once Received View that explanation involves showing *that* a phenomenon fits in the *nomic nexus*. Salmon's emphasis on causal relations over merely nomic relations was an important advance in

thinking about explanation. However, Salmon's final view of causation as an exchange of conserved quantities between intersecting causal processes is too narrow to encompass the diversity of causal relationships in neuroscience. Salmon's causal nexus does not (explicitly) recognize the importance of causal relationships at several distinct levels. And it does not adequately account for the notion of relevance that must lie at the heart of a mechanistic approach to explanation (see, e.g., Hitchcock 1995). If we are to make sense of the task of fitting a phenomenon into the causal nexus in neuroscience, we will have to extend our understanding of the causal nexus to make it into a properly mechanical nexus. In this section, I take some steps to work out some of the details of this extension.

2.1 Mechanisms. The mechanisms described in contemporary neuroscience are collections of entities and activities organized in the production of regular changes from start or set-up conditions to finish or termination conditions (See Machamer, Darden and Craver 2000; compare, e.g., Bechtel and Richardson 1993; Burian 1996; Glennan 1996; Salmon 1984; Thagard 2000; Wimsatt 1976). Mechanisms are how things work from start to finish, even if it is often a pragmatic matter where the mechanism starts and where it finishes (as in, for example, cyclic mechanisms). Mechanisms are composed of entities (what we typically call the "parts") with their various properties and of activities, the kinds of changing that make the mechanism work. There are many kinds of activities in contemporary neuroscience, including the generation of action potentials, the binding of neurotransmitters to a receptor, the processing of information or the back propagation of error, not all of which are usefully understood as exchanges of conserved quantities. Mechanisms do things; they are the mechanisms *of* the things that they do (like the

mechanisms *of* protein synthesis, or the mechanisms *of* synaptic facilitation). Mechanisms do what they do because they are *organized* spatially, temporally and interactively. The parts of the mechanism have sizes, locations, shapes, orientations, and containment relations, and the activities in mechanisms occur in particular orders, at particular rates and over various durations. The components are bound together into a single mechanism in part because of the causal interactions among them and in part because of their relevance (and the relevance of their interactions) to the behavior of the mechanism as a whole.

Consider one of the entities of contemporary neuroscience: the NMDA receptor-ionophore complex. This entity's name reflects what it does. It is an ionophore because it can bend itself into an ion channel traversing the membrane of the neuron. It is a receptor because it engages in this activity when certain neurotransmitters are present. One such chemical is N-methyl D-aspartate (NMDA), a pharmacological agent that mimics the endogenous neurotransmitter, glutamate.

Activation of the NMDA receptor is a means of transforming an extra-cellular chemical signal (born by neurotransmitters) and an intracellular electrical signal (born by ion fluxes in the cell) into an intracellular chemical signal (born by intracellular ions and molecules). The first two of these are the set-up conditions; the last is the termination condition. The extra-cellular chemical signal comes in the form of neurotransmitters (glutamate and glycine) that bind to extra-cellular binding sites. When the transmitters bind to the surface of the protein, the protein changes its conformation, exposing a channel through its center. Under resting electrical conditions of the postsynaptic cell, the ion channel is blocked by positively charged magnesium (Mg^{2+}) ions held in place by

electrical attraction and repulsion. This blockage is removed by the intracellular electrical signal— depolarization of the membrane. As the cell becomes less negative (and eventually positive) with respect to the extra-cellular fluid, the attractive and repulsive forces fixing the Mg^{2+} in the center of the channel weaken. The Mg^{2+} ions then drift out of the channel, allowing Ca^{2+} (the intracellular chemical "signal" constituting the termination condition) to diffuse into the cell.

This description includes the entities (e.g., glutamate, binding sites, channels, ions, membranes) and activities (e.g., binding, blocking, repelling, depolarizing) that intervene between the set-up and termination conditions. The components are organized spatially (e.g., the channel spans the membrane, the pore opens and closes, the extracellular Ca^{2+} moves into the cell), temporally (e.g., the depolarization precedes the release of the Mg+ ions) and interactively (e.g., the transmitter binds to the receptor, and binding changes its conformation). The organization of these parts gives rise to the behavior of the mechanism as a whole: a highly regulated gating of Ca^{2+} currents across the membrane.

2.2 Levels of Mechanisms. One reason neuroscientists are interested in the NMDA receptor is because its behavior is a component in the mechanism of Long Term Potentiation (LTP). LTP is one of the means by which certain neurons in the central nervous system (CNS) strengthen their connections (synapses) with one another. When the pre-synaptic neuron (the one that releases neurotransmitter) and the post-synaptic neuron (the one containing the NMDA receptor) are simultaneously active, the synapse is strengthened (LTP is "induced"). When the pre-synaptic neuron is active, it releases glutamate (and glycine) into the synaptic cleft. The post-synaptic cell is active when it is

depolarized from its resting electrical potential. These two factors, recall, are the crucial set-up conditions for the opening of the NMDA receptor. The termination condition (the influx of Ca^{2+}) is commonly believed to be a crucial stage in the induction of LTP. Many neuroscientists suspect that LTP is a crucial activity in the mechanisms of memory. According to one such hypothesis, LTP is a component in the mechanisms of "spatial map" formation in the hippocampus (a medial temporal lobe structure), and these spatial maps are thought to be components in the mechanisms of spatial memory, the ability to learn to navigate through novel environments.

This sketch of the theory of spatial memory describes a mechanism at multiple levels; it shares this feature with many theories in neuroscience and beyond. Zach Hall opens his *Introduction to Molecular Neuroscience* with the following paragraph:

> To understand how the nervous system works requires knowledge at several levels. At one extreme, we need to know how large numbers of neurons interact to produce the complex behavior of organisms. How do our eyes follow a moving object, for example? At another level, we need to know the properties of individual cells and how they interact. How do photoreceptors in the eye convey information about light to the rest of the nervous system? Finally, we want to understand the molecules of the nervous system and how they determine the properties and behavior of the cell. How do ion channels work? How do growing axons move? (1992)

In describing neuroscientific mechanisms as spanning multiple levels, we are using spatial metaphor. One aspect of this metaphor is *altitudinal*; we speak of levels as higher or lower than one another. Scientists often "go up a level" or "descend to" and "uncover" lower levels. Talk of levels and hierarchies also caries *compositional* commitments. Oppenheim and Putnam (1968) endorse this reading of the metaphor; one level is "lower" than another when the entities included in that level are proper parts of those at the higher

level. The NMDA receptor is part of the synapse, which is part of the hippocampus, which, in turn, is part of the spatial memory system.

Yet it is misleading (as Oppenheim and Putnam acknowledge) to focus exclusively on mereological relationships among entities and their parts. Exclusive focus on mereological relations obscures the extent to which the levels in the theories of neuroscience are shaped by (in addition to part-whole relations) the activities in which those entities engage. In the above example, the NMDA receptor is of interest because the entities the activities in the mechanism of activation are those that are *relevant* for understanding how the channel opens under those set-up conditions.[3] Likewise, this capacity of the NMDA receptor is of interest by virtue of the fact that it is a stage in the mechanisms of LTP induction, which constitutes a stage in the mechanism of spatial map formation, and so on. There are many ways of breaking things into parts; they can be sliced, diced, cubed or spiral cut. But when we break a mechanism into working parts, what will count as a part and how the parts will be individuated can only be assessed by reference to the relevance of these parts to the behavior of the mechanism as a whole. The same holds for aspects of the mechanism's spatial, temporal and interactive organization. Brute part-whole relationships do not adequately address the relevance of the parts to the behavior of the mechanism as a whole or to the ways that they are organized together to produce that behavior.

For this reason, the hierarchies of contemporary neuroscience are usefully thought of as hierarchies of mechanisms: they are part-whole relations with the additional

[3] I do not here develop this notion of relevance. But note that even if exchanges of conserved quantities could establish relevance in cases of etiological causation, they are not made to handle the interlevel relevance of parts to the behavior of the mechanism as a whole. There is no exchange of conserved quantities in interlevel mechanistic (i.e., constitutive) causation.

restriction that the parts are organized together to produce the behavior of the mechanism as a whole. To be at a lower (-1) level is just to be one of the components organized into the mechanism as a whole, which constitutes the higher (+1) level. Of course, there are other useful notions of "level" (tracking e.g., objects of different sizes, or the phenomena in different theories, or the domains of different sciences). However, levels of mechanisms capture a common notion of level that is well-represented in the theories of contemporary neuroscience and that captures what Hall describes in his introduction. (It is, incidentally, also one that sometimes does correspond to differences in size, among theories and among scientific fields).

Hierarchies of mechanisms *bottom out* and *top off*. Descriptions of mechanisms bottom out when further redescription in terms of component entities and activities is unnecessary, unavailable or inappropriate. Different scientists, fields and traditions bottom out their hierarchies in different mechanisms. For some neuroscientists, the above description of the NMDA receptor bottoms out; others anxiously await an understanding of the mechanisms by which proteins change their conformation. This relativity of bottoming out also holds for the level at which these hierarchical descriptions top off—the upper-most level in a hierarchy of mechanisms. The choice of this top level selectively focuses attention upon certain lower-level mechanisms and not others. Different scientists top off in different highest-level activities: e.g., the survival and/or reproduction of the organism, the production of a protein, the progression of a disease, or the stability of an ecosystem. Differences in topping off points reflect differences in interest and emphasis (these are, in fact, one kind of Wimsattian perspective), and these differences are reflected in the theories that different neuroscientists, fields or traditions

use to explain the phenomena in their domain. It is by reference to this historically and disciplinarily relative topping off point that the relevance of lower-level components is fixed.

In our working example, it is an antecedent interest in spatial memory that focuses the investigator's attention upon the mechanisms of spatial map formation, LTP, and the mechanisms of NMDA receptor activation. The mechanisms of NMDA receptor activation are also included in hierarchies that top off in the progression of Huntington's and Alzheimer's disease, the psychological effects of PCP abuse, the mechanisms of programmed cell death and locomotion. Differences in topping off point are typically reflected in different mechanisms to explain them and different levels between top and bottom.

Descriptions of mechanisms and the hierarchies that they compose provide the conceptual structure within which the neuroscientist explains the workings of the CNS. Such explanations, as Salmon suggested, involve situating the item to be explained (an entity, property, activity or type thereof) within the nexus of mechanisms. This process of "situating" an item within the causal nexus is a matter of tracing a pattern or pathway in the nexus by trimming away irrelevancies and relating what remains to the item to be explained. This may be done in at least three ways.

3. Etiological Explanation and Adaptational Functions. Perhaps the most familiar variety of causal mechanical explanation involves identifying antecedent, or efficient causes. Etiological explanations are typically offered in response to questions concerning the origins of some item, its path of development, or its historical trajectory. For example, we seek etiological explanations for the existence of life on the planet, for

the course of an epidemic, or for the events leading up to the declaration of war. We also seek etiological explanations for the presence of certain traits in biological populations. A request for etiological explanation is answered by identifying some set of start-conditions in the item's past and tracing the relevant entities and activities intervening between those conditions and the item to be explained (treating it as a termination condition). Salmon represents this "etiological aspect" of mechanistic explanation in the bottom portion of Figure 1. The figure illustrates the *backward looking* character of etiological explanations; such explanations highlight the pathway connecting relevant set-up conditions in the past, through intermediate stages of activity, to the item to be explained.

[Figure 1 Near Here]

There are a number of philosophers and scientists alike who reserve the term "function" for traits, properties, and activities that are adaptations (Ruse 1971; Brandon 1990). Churchland and Sejnowski believe that this use adequately captures the sense of "function" used in neuroscience (1992; 69).[4] Variant analyses of this adaptational sense of "function" have been developed by Millikan (1984; 1989), Neander (1991), Wimsatt (1972) and Wright (1973). On Wright's classic formulation:

> The function of X is Z means:
>
> (a) X is there because it does Z,
>
> (b) Z is a consequence of X's being there (1973, 161).

In the standard biological case, (a) is embellished as a natural-selection story roughly to the effect that heritable traits of type X, by virtue of their doing Z, made those organisms

[4] Their discussion of this issue, however, is rather noncommittal. They note that the function of an item is its "job," that any apparent teleology in the sense of function is "eliminable or reducible without remainder in an evolutionary framework," and that functions "help the animal move adaptively in the world" (69). As

bearing traits of type X more likely to survive and/or reproduce, and as a result helped to preserve traits of type X in a given population of organisms. The mouse's NMDA receptor has the adaptational function of mediating cellular "signals" if and only if the NMDA receptor allows these signals to be mediated within the mouse and was preserved in the population of mice because it did so in the past. One advantage of identifying biological functions with adaptations is that doing so can often accommodate the intuition that a trait's function explains its presence. Asked why the mouse has NMDA receptors, it may be correct to respond that the NMDA receptors are there *because* they mediate certain chemical signals and that doing so is step in LTP induction, spatial map formation, and spatial memory. Still, adaptational explanation is straightforwardly mechanistic; it is an example of the *etiological* type of mechanistic explanation.

There are a number of motivations for thinking of functional ascriptions as condensed and elliptical adaptational (and so etiological) explanations. This restrictive view has been voiced by Bechtel 1989, Lycan 1987 and Neander 1991, among others. I do not think that functional ascription in the neurosciences should not be understood exclusively, or even primarily, in this manner. There are philosophical motivations for forging such associations, but these are quite alien from the explanatory concerns of the neuroscientist.

First, functional language appears in descriptions of mechanisms that in no way aid in survival and reproduction. To assume otherwise is to neglect the fact that a great deal of neuroscience is driven not by the goal of understanding how the nervous system functions when it is working properly but rather by the goal of understanding how it can

---

the following discussion will make clear, these remarks (save the second) are consistent with a number of

fail and how those failures might be predicted and controlled. Mutations, binding cites for pharmacological agents, characteristic anatomical markers of psychiatric disorders: these are as much a part of the mechanisms of the brain as those that evolution left there for a reason. There are many ways of defining the topping-off points for theories, depending on the end that the theory is intended to serve: compare the teleofunctionalist perspective with that of those approaching the brain from the perspectives of agricultural production, behavior management, product design, and military applications. It is simply a mistake to assume that all of neuroscience (or biology) is concerned with understanding how individuals and species survive and reproduce, or to assume that functional language is appropriate only for describing phenomena relevant to such ends.[5]

Yet even for the central cases in which an item can be shown to increase the probability that an organism of some type will survive and reproduce, the neuroscientist is willing to describe it functionally while remaining agnostic with respect to its origins. Although neuroscientists, in their unguarded moments, often speak *as if* the CNS and its components were designed (even *optimally*) through the process of evolution by natural selection, their public timidity with such locution reflects an underlying recognition that matters are not so simple. They understand, for instance, that the trait may have been preserved in the population as a result of genetic drift (or repeated inbreeding of laboratory animals), or as a "piggyback" on another trait, or that it might have been recruited for a new function (i.e., exapted) after having been previously selected for another (Gould and Vrba 1982; Gould and Lewontin 1978). Such complicated etiologies

---

different approaches to the concept of function.

[5] Analyses of the concept of function in terms of current ability or propensity to survive and reproduce (e.g., Bigelow and Pargetter 1987; Boorse 1976; Canfield 1964) likewise fail to accommodate many of the perfectly legitimate uses of functional language that can be found in neuroscience.

are, in fact, expected on the basis of the known malleability of the brain and the complexities of brain development.

Furthermore, our understanding of the brain's evolution (and the evolution of its constituents) is limited. If we were forced to appeal to evolutionary or etiological factors to assign functions, we would be operating largely in the dark. Behaviors and soft tissues do not survive in the fossil record. Comparative anatomy and studies of development have provided some clues, but there are still rather gaping holes in our etiological explanation of the details of brain structure and function. For example, John Eccles, in *Evolution of the Brain: Creation of the Self* (1989), was reduced to a sort of phrenological evidence common in the prehistory of neuroscience: comparing the sizes of different brain regions across extent species and cranial volumes across extinct species. Such indices, while potentially revealing, are hardly scalpels for understanding the functional organization of the brain. And interpretation of such volumetric indices rests upon a *prior* understanding of the function of the brain regions whose volumes are being compared. The evidence upon which neuroscientists ground their functional attributions is typically found in the laboratory or the clinic, not in the evolutionary lineage. (See Section 3.3).

The adaptational sense of function is useful as a terse indicator that a trait has a particular sort of etiological explanation. But there are a number of seemingly legitimate uses of functional description that fall outside its purview. Sometimes an item's presence can be explained by its function, and sometimes it cannot.

4. **Constitutive Explanations and IO Functions.** There is a common use of the notion of "function" in contemporary neuroscience according to which a function is a

mapping from inputs to outputs in conformity with a rule. Call these "input-output" (IO) functions. Sometimes Cummins describes functions this way. For example, he says that functions are capacities and that capacities are, "specified by giving a special law linking precipitating conditions to manifestations— i.e., by specifying input-output conditions" (1983, 53). Turning machine (TM) functionalists can be read as endorsing this view. IO functions characterize the activity of some item without reference either to its context or to its internal complexities. In forming such a description, one draws a conceptual dividing line at the spatial boundary of the object and recognizes a limited number of specific *interfaces* across that boundary—more or less well-defined interactions with items outside of that boundary. For example, in describing the IO function of the NMDA receptor, itself a complex of a number of entities, one begins by parsing it from its environment at its spatial boundary and characterizing the relevant interfaces (interactions) across that boundary. Three significant interfaces between the NMDA receptor and its environment are the binding of glutamate and glycine to the receptor, the blocking action of Mg^{2+} ions, and the influx of Ca^{2+} through the channel pore.

[Insert Figures 2 a and b Near Here]

IO functions obtaining among these interfaces (and others like them) can sometimes be characterized using equations and generalizations. Two such equations are represented graphically in Figures 2a and 2b (taken from Mayer et al. 1991). The first figure depicts a typical dose-response curve relating concentrations of aonists (glutamate and pharmacological concoctions) to the current of Ca^{2+} flowing into the post-synaptic cell in the absence of Mg^{2+}. The second of these characterizes changes in that influx of Ca^{2+} as a function of post-synaptic depolarization in a medium with extremely high

concentrations of Mg^{2+}. Clearly neither of these fully characterizes the activation of the NMDA receptor (i.e., its function), and each characterizes it only under highly constrained conditions (i.e., experimentally gerrymandered levels of Mg^{2+}). Rather, these IO functions and the myriad others like them combine to form a complex description of the behavior of the NMDA receptor. (Note again that which complex IO function for an item one takes as relevant will depend upon the choice of mechanism in which one includes the item. The IO function for the NMDA receptor may look different for a scientist interested in the mechanisms of learning and memory than it does to those interested in programmed cell death.

This *complex IO function* plays two important roles for the neuroscientist. First, describing an item functionally, it affords necessary descriptive leverage over the messy details of the constitutive mechanism (or type of mechanism) producing the complex IO function. One can speak of the activation of the NMDA receptor without going into the messy of protein chemistry, or can speak of LTP induction without detailing the intricate pattern of molecular activities responsible for it. IO functions also offer a descriptive tool for dealing with individual, strain and species differences, since the same IO function might be instantiated by a number of different mechanisms (although biological diversity provides reason to doubt that this is as common as TM functionalists would believe). They are also useful for dealing with the fact that individual mechanisms can themselves change over time to perform the same function in different ways (e.g., learning, development, or reorganization due to injury). In short, the IO function glosses over this detail and diversity in the mechanisms at the lower level.

Complex IO functions are also important for framing *constitutive explanations.* The term is adapted from Salmon's (1984) discussion of the "constitutive aspect" of causal mechanical explanation. The term is intended to capture the fact that such explanations proceed by describing the internal mechanisms— organized lower (-1) level activities and entities— by virtue of which some aspect of the complex IO function is produced. This type of explanation is represented by the encircled volume of space-time in the center of Figure 1.[6] Constitutive explanations are sought when one wants to know how something works or wants to know the "hidden" mechanism by virtue of which an item does something of interest. The explanation of the opening of the NMDA receptor in Section 2 is an example of this constitutive form of mechanistic explanation. That explanation is tailored to account for the IO functions represented in Figures 2a and 2b and the myriad others like them. It is in this sense that the complex IO function frames the constitutive explanation: it defines the relevant input-output relationships that the internal mechanism must be capable of performing. Whereas etiological explanations situate an item in the causal nexus by "looking backwards," constitutive explanations are *downward looking*; they situate an item in the causal nexus by detailing the lower-level mechanism producing those aspects of the complex IO function.

To preface a contrast underlying Section 6: IO functions are "narrow." They describe the behavior of an item cut off from its context in the causal nexus. Descriptions that include this context describe the role of the item within that contextual mechanism. The same IO function can serve different roles depending upon the context in which it is

[6] Cummins (1975) uses the term "functional explanation" for the decomposition of an analyzed capacity into component capacities; the terms "mechanistic explanation" or "homuncular explanation" seem to me each to be prefereable. Cummins idiom is confusing, given that functional decomposition is typically

situated; the same role can be served by different IO functions; and one can know an item's role without knowing its IO function, and vice versa.

One cannot describe an item's role, in the broad sense intended here, without describing the place of its IO function in some more inclusive mechanism. As noted above, the capacity of the NMDA receptor to open under appropriate set-up conditions has different roles depending upon whether it is contextualized within the mechanism responsible for LTP induction, Huntington's disease, or programmed cell death. In one case, the NMDA receptor is the post-synaptic "gateway" to the production of LTP; in another, it is one of the entities implicated in neuronal death. What has changed is not the NMDA receptor, its IO function, or its activities (although different aspects of those descriptions may become more or less salient from these different perspectives). What has changed is its context, and with that change comes a change in roles.

Moving to a higher level in the spatial memory hierarchy, the same IO function of the hippocampus has a role in a variety of memory functions (in humans, it is associated with "declarative memory" or "memory of facts"). To interpret the IO function of the hippocampus as "forming a spatial map" involves implicitly embedding that IO function in the context of various brain regions responsible for spatial memory in particular, and also among the features of the world which are relevant to spatial navigation. Currently, researchers have a working understanding of the various roles of the hippocampus in the spatial memory system although there is no consensus as to the IO functions by which it performs those roles. This suggests a pragmatic need for keeping roles and IO functions distinct.

---

associated with describing mechanisms whereas functional explanations are typically concerned with

Functions as roles are not simply capacities *picked out* by their place in a higher level mechanism (as Cummins once suggested); rather, they are descriptions of the activity of some item explicitly in terms of how it is organized into the workings of a higher (+1) level mechanism. One and the same token sparking of a spark plug may be said to be an instance of sparking, of igniting an explosion, of pushing a piston, and of turning the drive shaft depending on how much of the item's context in the causal nexus one includes in the description.

5. Contextual Explanation and Role Functions. Functional descriptions in neuroscience often (even typically) include extensive description of the way that an item is situated in its causal context. A classic example is the concept of a neurotransmitter. What evidence does it take to establish that some chemical substance is a neurotransmitter?

[Table 1 Near Here]

Table 1 lists six criteria that appear in most introductory neuroscience texts. Although many of these criteria are violated for known neurotransmitters (especially amino acid transmitters like glutamate), they are nonetheless give a picture of the kinds of evidence relevant to evaluating a crisp and precise functional description. Criteria 1-6 are designed to show that the putative neurotransmitter is situated within the mechanisms of chemical neurotransmission. These mechanisms are represented in Figure 3. The first criterion for identifying a neurotransmitter is *that the transmitter must be present in the axon terminal*. In Figure 3, the transmitter is stored within circular vesicles floating in the cytoplasm of the axon terminal (see item C). The second criterion is that *the substance*

---

looking upward, as we will see (See also Bechtel and Richardson 1993; Wimsatt 1972; Burian 1996).

should be released in a calcium- and depolarization- dependent manner, and should be released in amounts sufficient to exert its supposed action on the post-synaptic cell. The release of the substance should be calcium and depolarization dependent because the constitutive explanation of transmitter release makes reference to depolarization-induced opening of voltage-sensitive Ca^{2+} channels in the axon terminal (See Figure 3; items 1-6). On the post-synaptic side, if a substance is to be considered a chemical messenger between two cells, then there should be some post-synaptic response to that transmitter (See Figure 3; items 10-12). The third criterion requires that these post-synaptic *effects be produced by both exogenous and endogenous application of the chemical*. Evidence concerning pre-synaptic synthesis, the fourth criterion, is required to show that the chemical's production is subject to a form of regulation typical of neurotransmitters. This criterion is clearly designed to demonstrate the chemical's active organization within the mechanisms of synaptic signaling. The fifth criterion, that there is some mechanism for removing neurotransmitter from the cleft, is required if the concentration of the cleft is to correlate with the activity of the pre-synaptic cell. The tight relationship between action potentials in the pre-synaptic cell and transmitter concentration in the cleft would not hold were there no mechanisms for disposing of "excess" transmitter in a timely fashion (See Figure 3; items 7, 8, and 8a). Finally, *the putative post-synaptic effect of the chemical substance should be mimicked or blocked by pharmacological agents known to activate or impede the post-synaptic receptors for that substance*. Neurotransmitters transmit by binding to receptors on the post-synaptic cell. Mimicking or preventing that binding should change the post-synaptic response.

So what makes a neurotransmitter a neurotransmitter? Two possible answers can be ruled out from the start.

The first is the chemical's pedigree. For Wright (Section 3) to claim that a substance has the function of mediating communication between cells (as evidenced by the six criteria) would involve showing (i) that the chemical substance is capable of mediating communication between cells, and (ii) that the chemical substance came to be at this synapse because it can mediate communication between cells. Although criteria (1-6) are clearly designed to satisfy some requirement like (i), none of them address (ii). Rather, the evidence for (i) is our best evidence for believing that (ii) is true. Regardless of how the molecule came to be used as a neurotransmitter (by drift, exaptation, chance or divine fiat), it still functions as a neurotransmitter at the synapse.

The second answer that can be ruled out on the basis of this example is that functioning as a neurotransmitter is an IO function. What is at stake is not a local set of interactions at the surface of the molecular transmitter. Instead, these six criteria are designed to show that a given chemical substance is situated among the mechanisms of chemical transmission in such a way that it can fulfill the role of a neurotransmitter. This involves not merely specifying some IO function of the chemical substance (although criteria 3 and 6 do address IO functions). Instead, it involves showing that the exercise of these capacities is tightly woven into the mechanisms for regulating the chemical's synthesis, release and removal from the synapse. Describing the isolated interactions of the molecule is like describing the sparking of the spark plug.

Here we have a third sense of function— role functions— which answer to a particular sort of explanatory need in sciences whose subject exhibits the hierarchical

structure discussed in Section 2. This explanatory need is fulfilled by a third form of mechanistic explanation, which I call *contextual explanation*. Often, the neuroscientist is ignorant of the role of some item, and this leads her to ask after the role of some item in a mechanism. The answer to such a request for explanation comes in the form of a description of how an item is organized together with the other components in a higher level mechanism. The process of situating an item in a higher (+1) level mechanism involves showing how it is organized (spatially, temporally and actively) into the higher level mechanism[7]. The neurotransmitter has to be released in correlation with the electrical properties of the cell, has to be cleared from the cleft, has to act on post-synaptic receptors, and has to exhibit the kinds of active organization within a mechanism revealed by the other criteria in Table 1. The concept of a neurotransmitter, as one of our well-articulated concepts in contemporary neuroscience, provides a model of a contentful functional ascription and of the kinds of evidence by which such ascriptions are to be evaluated.

Consider now an example of a disconfirmed functional description. Prior to the 1950's the hippocampus was generally considered to be an organ of olfaction. Various lines of evidence supported this conjecture: the hippocampus was located in the rhinencephalon, patients with temporal lobe epilepsy often reported olfactory auras, and extirpation of the hippocampus had been noted to produce profound deficits in the ability of dogs and other animals to perform olfactory discrimination tasks after the surgery. In 1941, Alf Brodal could correctly proclaim that nearly every textbook described the

[7] The sort of situatedness intended here has analogues in explanation in history. One can, for example, ask after the significance of the shot at Serejevo or Darwin's Beagle voyage. Answers to such questions trace the historical context (antecedent, contemporaneous, and subsequent) of these historical occurrences and

hippocampus as an olfactory organ. Contrary evidence from lesion experiments was also available at the time, but the most decisive evidence came in the form of a negative result: the hippocampus could not be shown to have any anatomical connections with the olfactory bulb. The functional hypothesis failed because the hippocampus was demonstrated not to be organized with other components in such a way as to have a role in the olfactory system.

If one accepts Salmon's broad outline of causal mechanical explanations (i.e., that they explain by showing how an item or event fits into a nexus of causes), then there is a clear candidate for a third aspect or variety of causal mechanical explanation: a contextual explanation. A contextual explanation explains an entity or activity by showing what it is for, that is, how it fits into the organization of a higher-level mechanism. Contextual explanations are characteristically outward looking and upward looking. They are outward looking because they refer to components outside of the item to be explained and they are upward looking because they contextualize that item within the behaviors of a higher (+1) level mechanism. Contextual explanations show quite literally how an entity or activity fits into a mechanism. So there are not two, but three ways of showing how something fits into the causal nexus.

This perspective on functional attribution suggests a regulative ideal in formulating functional attributions: they are contentful and precise to the extent that they explicitly make claims about how an item is situated in its causal context. It is by reference to the evidence for such organization (as is the case for neurotransmitters) that role attributions are evaluated.

---

show how the explanandum fits into an historical sequence that is antecedently assumed to be interesting or

Opponents of "analytic accounts" of functions will object that this contextual notion of function is too liberal: a given item is likely to be situated within the causal nexus in innumerable ways. Without some further restriction on which containing systems count in the attributions of functions, we end up countenancing innumerable functions: it is the function of the hippocampus to use glucose, the function of the heart to make glub-blub noises, or the function of muscles to produce heat. We are now in a position, however, to see why this problem should not trouble us. Van Frassen, in his criticisms of Salmon's causal-mechanical perspective, argued that our assessments of etiological causes are laden with pragmatic concerns: was it the alcohol, the ice on the road, or the intrusive thoughts of Marilyn and her lover that caused the car leave the California highway? There are usually a number of correct (if partial) ways of situating something with respect to the past portions of the causal nexus. Given that componency in mechanisms is established by reference to some behavior that one seeks to understand (e.g., spatial memory, apoptosis, or Alzheimer's diseas), we have also seen that constitutive mechanisms are tinged with the same pragmatic concerns. It would be surprising if now contextual mechanisms did not share in this pragmatic aspect. There is a fact of the matter about the causal connections composing an item's contextual causal nexus, but there is no uniquely correct fact of the matter as to which of those connections in the causal nexus ought to be highlighted in providing a response to a particular request for explanation. This is not a spooky variety of interest relativity, but the simple recognition that some aspects of the causal structure of the world are more interesting and relevant to our purposes than others. That should be explicit in our account of how we

---

important.

assign functions to the brain and its parts, it should not masked with obfuscating, vague and evidentially tenuous talk about contributions to fitness and propensities to survive or reproduce.

[Figure 4 Near Here]

6. Conclusion. Three explanatory perspectives are illustrated in Figure 4. The figure depicts two levels (the top and bottom circles) in a mechanistic hierarchy flanking the complex IO function in the middle. The past and future portions of the causal nexus are to the left and right of the hierarchy, respectively. For each type of explanation, the explanandum is some aspect of the complex IO function in the middle; call it E.

An etiological explanation traces the pathway of entities and activities terminating in E; they explain how E came to be there, came to pass, or came to have some property. Explanation in terms of natural selection is a type of etiological explanation, one in which items of the same type as E appear in the explanans. Citing E's adaptational function is an important part of giving its etiological explanation. Adaptational explanations are thus *backward looking*; they are also legitimate answers to a certain reading of the question "Why is E there?"

Constitutive explanations explain how E works. They are *downward looking* in that they situate E with respect to the portion of the causal nexus at a lower (-1) level in a hierarchy of mechanisms. E is a "black box," but if we look within, we find that it is composed of the pattern of entities and activities at that (-1) level. Complex IO functions are especially useful for describing E without reference to such messy details, but they also frame internal (-1) mechanistic explanations; it is a requirement on the adequacy of

such explanations that they account (more or less) for the input-output functions of the mechanism as a whole.

Finally, contextual explanations are *upward looking*; they situate E with respect to the portion of the causal nexus in the higher (+1) level of a hierarchy of mechanisms. This is why it is explanatory to cite E's role; role-descriptions provide a more or less terse description of how E is related to the other entities and activities in a (+1) level mechanism. They are therefore legitimate answers to a second reading of the question, "Why is E there?"

Mechanistic description and functional description are equally widespread in contemporary neuroscience. This is not a coincidence, since functional ascriptions (usually ascriptions of contextual roles) are important determinants of the structure of neuroscientific theories. Assigning functions to an item integrates that item into its mechanistic context. It is for this reason that functional descriptions are precise and contentful to the extent that they embody commitments as to how those contextual mechanisms are organized.

REFERENCES

Bechtel, W. 1989. "Teleological Functional Analyses and the Hierarchical Organization of Nature", in N. rescher (ed.), Teleology and Natural Science. Landham, MD: University Press of America, 26-48.

Bechtel, W., and Richardson, R. 1993. *Discovering Complexity*. Princeton: Princeton University Press.

Bigelow, R. and Pargetter, R. 1987. "Functions." *Journal of Philosophy* 84:181-97.

Boorse, C. 1976. "Wright on Functions." in Sober E., 1984, 369-85.

Burian, R. M. 1996. "Underappreciated Pathways Toward Molecular Genetics as

Illustrated by Jean Brachet's Cytochemical Embryology," in ed. S. Sarkar, *The Philosophy and History of Molecular Biology: New Perspectives*. Dordrecht: Kluwer. 67-85.

Canfield, J. 1964. "Teleological Explanation in Biology." *British Journal for Philosophy of Science* 14: 285-95.

Churchland, P. S., and Sejnowski, T.J. 1992. *The Computational Brain*. Cambridge, Mass.: MIT Press.

Cummins, R. 1975. "Functional Analysis." *Journal of Philosophy 72*: 741-64. Reprinted in Sober, E., 1984, 386-407.

___. 1983. *The Nature of Psychological Explanation*. Cambridge, MA: Bradford/MIT Press.

Craver, C.F. (2001) "Role Functions, Mechanism, and Hierarchy." Philosophy of Science 68: 53-74.

Eccles, J.C. 1989. *Evolution of the Brain: Creation of the Self*. Routledge: London.

Glennan, S.S. 1996. "Mechanisms and the Nature of Causation." *Erkenntnis* 44: 49-71.

Gould, S.J. and Lewontin, R.C. 1978. "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme." *Proceedings of the Royal Society of London* 205: 581-598.

Gould, S.J. and Vrba, E. 1982. "Exaptation – A Missing Term in the Science of Form." *Paleobiology* 8: 4-15.

Hall, Zach 1992. An *Introduction to Molecular Neurobiology*. Sunderland, MA: Sinauer.

Hitchcock, C. R. 1995. "Discussion: Salmon on Explanatory Relevance," *Philosophy of Science* 62: 304-320.

Lycan, W. 1987. *Consciousness*. Bradford Books/MIT Press.

Machamer, P.K., Darden, L., and Craver, C. F. 2000. "Thinking About Mechanisms." *Philosophy of Science* 67: 1-25.

Mayer, et al. 1991. "NMDA Receptors: Physiological Studies with Divalent Cations and Competitive Antagonists." In *Neuroscience of the NMDA Receptor*, ed. A. Kosikowski and G. Barinuevo, New York: VCH Publishers.15-35.

Millikan, R. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism.* Caimbridge, MA: MIT Press.

Neander, K. 1991. “Functions as Selected Effects: The Conceptual Analysts Defense.” *Philosophy of Science* 58:168-84.

Oppenheim, P. and Putnam, H. 1968. “Unity of Science as a Working Hypothesis.” In *Concepts, Theories, and the Mind-Body Problem*, *Minnesota Studies in the Philosophy of Science II*, eds. H. Feigl, M. Scriven, and G. Maxwell, 3-36. Minneapolis: University of Minnesota Press.

Rosenberg, A. 1985. *The Structure of Biological Science*. Cambridge: Cambridge University Press.

Salmon, W.C., 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Shepherd, G.M. 1994. *Neurobiology*. Oxford: Oxford University Press.

Thagard, P. 2000. “Explaining Disease: Correlations, Causes and Mechanisms” in R. Keil and R. Wilson eds. Explanation and Cognition. Cambridge, MA: MIT Press.

Wimsatt, W. C. 1972. “Teleology and the Logical Structure of Function Statements.” *Studies in History and Philosophy of Science* 3: 1-80.

___. 1976. “Reductive Explanation: A Functional Account.” In R.S. Cohen et al., eds. Reprinted in Sober, 1984, 477-508.

Wouters, A.G. 1999. *Explanation Without a Cause*. Amsterdam: Zeno Institute.

Wright, L. 1973. “Functions.” *Philosophical Review* 85: 70-86. Reprinted in Sober, 1984, 347-68.