Garson, J. (2012) "Function, Selection, and Construction in the Brain," *Synthese* 189: 451-481.
Preprint (not copyedited or formatted)
Please use DOI when citing or quoting: http://dx.doi.org/10.1007/s11229-012-0122-y

Title: Function, Selection, and Construction in the Brain

Abstract A common misunderstanding of the selected effects theory of function is that natural selection operating over an evolutionary time scale is the only function-bestowing process in the natural world. This construal of the selected effects theory conflicts with the existence and ubiquity of neurobiological functions that are evolutionary novel, such as structures underlying reading ability. This conflict has suggested to some that, while the selected effects theory may be relevant to some areas of evolutionary biology, its relevance to neuroscience is marginal. This line of reasoning, however, neglects the fact that synapses, entire neurons, and potentially groups of neurons can undergo a type of selection analogous to natural selection operating over an evolutionary time scale. In the following, I argue that *neural* selection should be construed, by the selected effect theorist, as a distinct type of function-bestowing process in addition to natural selection. After explicating a generalized selected effects theory of function and distinguishing it from similar attempts to extend the selected effects theory, I do four things. First, I show how it allows one to identify neural selection as a distinct function-bestowing process, in contrast to other forms of neural structure formation such as neural construction. Second, I defend the view from one major criticism, and in so doing I clarify the content of the view. Third, I examine drug addiction to show the potential relevance of neural selection to neuroscientific and psychological research. Finally, I endorse a modest pluralism of function concepts within biology.

1 Brain functions result from selection processes

According to the selected effects theory of function, the function of a biological trait consists (roughly) in the effect that it was selected for by a natural process of selection (e.g., Neander 2008, 386; also see Neander 1983; 1991; Millikan 1984; 1989; Sober 1984; Brandon 1990; Griffiths 1993; Godfrey-Smith 1994; Mitchell 1995; Allen and Bekoff 1995; Schwartz 1999). This theory has often been coupled with an auxiliary assumption, namely, that the only selection process relevant for the production of functions is natural selection operating over an evolutionary time scale and paradigmatically at the level of the individual organism (e.g., Sober 1984, 208; Brandon 1990, 186; Neander 1991, 174; Allen and Bekoff 1995, 612; Walsh and Ariew 1996, 497; Wouters 2003, 649-652; Lewens 2007, 533).

This secondary assumption creates two important obstacles for the empirically-warranted ascription of functions in the context of neuroscience (e.g., Craver forthcoming; Garson 2011). The first is that the evolutionary history of many features of the brain that neuroscientists are interested in is largely unknown and remains shrouded in speculation, such as long-term changes in the evolutionary function of neuropeptides oxytocin and vasopressin (e.g., Porges and Carter 2011). All things being equal, a theory of function that provides a cogent means by which to identify empirically the function of a trait is preferable to a theory of function that undermines severely the epistemic warrant for function ascriptions. The second problem is the ubiquity of evolutionarily novel brain functions, such as the function of the visual word form area (VWFA) in facilitating

reading ability (Schlaggar and McCandliss 2007) or the function of certain areas of the temporal lobe in facilitating Tetris play (Haier et al. 2009), which could not have been selected for over an evolutionary time scale. These kinds of problems have been used to suggest that, though the selected effects theory may be useful (at best) in certain areas of evolutionary biology, it is not applicable in most other areas of biology (Griffiths 2006, 3) and that some other theory, for example, the causal role theory, is more appropriate in those contexts (Craver forthcoming).

Selection processes, however, are ubiquitous in the natural world. In addition to natural selection operating over an evolutionary time scale, there are also neural selection processes, antibody selection processes, and some types of learning that can be modeled as selection processes, such as instrumental learning (operant conditioning). These latter three processes can all be construed as selection processes in that they lead to the *differential* reproduction, replication, or reinforcement of certain entities over others in a manner that is formally analogous to the process of natural selection operating over a population of reproducing organisms (Darden and Cain 1989; Cziko 1995; Hull et al. 2001). This observation is not novel in the functions literature; some function theorists have explicitly claimed that there are different kinds of selection processes in the natural world and each can give rise to novel functions (Wimsatt 1972; Millikan 1984; Papineau 1987, 1993; Godfrey-Smith 1992). These attempts have informed my preferred definition of "function," which I call the *generalized* selected effects theory of function (GSE). According to GSE, the function of a trait consists in that effect that contributed to the

differential reproduction *or* differential retention of that trait within a biological population (Garson 2011; also see Garson 2010, 222-225).

The purpose of the following is to articulate carefully how GSE applies to neural selection. In doing so it shows how neural selection processes may explain how novel functions emerge on an ongoing basis over the lifetime of the individual and not merely over an evolutionary time scale. I do not claim that all selected effects functions in the brain result from neural selection; some of them result from natural selection (for example, by increasing the population-wide frequency of "genes for" distinct neural pathways). There may even be conflicts between the functions that emerge from different selection processes; a structure that would be considered dysfunctional from the reference point of natural selection may be functional from the reference point of neural selection. Drug addiction may be an example (see Section 5).

As a consequence of these considerations, the selected effects theory emerges as an important complement to other theories of function that have dominated philosophical reflection on neuroscience, in particular, the causal role view (e.g., Bechtel and Richardson 1993; Craver 2001, forthcoming; Glennan 2005, 456; Piccinini and Craver 2011; Weiskopf 2011; Kaplan 2011). As will be developed in the concluding section, my view is that etiological function theorists unnecessarily "ceded" the context of neuroscience to causal role theorists because of their misplaced emphasis on *evolutionary* considerations. This led some theorists to reason that, since the selected effects theory only deals with evolutionary questions, it is not relevant to the context of neuroscience,

one primary concern of which is the functional decomposition of complex abilities independent of evolutionary history. Griffiths (2006, 3), for example, argues that the selected effects theory of function is relevant only to certain sub-branches of evolutionary biology; outside of evolutionary biology, the causal role theory reigns supreme. In the final section I argue that this view is mistaken, because the selected effects theory addresses a broader range of questions than merely evolutionary ones. Moreover, considering the selected effects functions of various neural structures may allow one to make novel predictions about the current-day causal roles of those structures, and to design rational methods of biomedical and psychiatric treatment. This is not to defend a functional monism, but rather, to insist on the utility of functional pluralism in many different branches of biology.

In addition to its importance for the functions debate, there are a number of reasons why reflection on neural selection may be philosophically rewarding. One reason is that it provides an important touchstone for philosophical explications of the concept of natural selection itself, though it has been largely neglected in the literature on natural selection in the philosophy of biology (but see Darden and Cain 1989). Second, it may contribute constructively to teleosemantic theories of mental content. According to teleosemantics, the content of a mental representation is determined, in part, by its selection history (or the selection history of its constituent representations). One problem with this view is that, while natural selection may explain the content of rather crude representational states, such as *fire* or *predator*, it doesn"t seem to explain the contents of representations that aren"t directly "exposed" to natural selection, such as *The Communist Manifesto* or

flatbed auto-feed scanner. Thus, teleosemantic theories seem committed to some distinction between *simple* and *complex* representations, where simple representations derive their contents from their selection history and complex representations derive their contents by virtue of the way that they combine simple representations (Neander 1999, 22). If *neural selection* processes can confer contents upon certain mental representations, however, then it can produce novel *simple* representations. As a consequence, the set of *simple* representations would be much larger than the set of simple representations that derive their content from natural selection. This would provide the teleosemantic theorist with a larger number of tools for explaining the contents of diverse representations.

Finally, neural selection has played an important role in recent criticisms of evolutionary psychology, and in particular, the massive modularity thesis. Thus, reflecting on neural selection may inform discussions about the architecture of the mind. Buller (2005) and Buller and Hardcastle (2000) argue that the cognitive capacities studied by evolutionary psychologists are probably not genetically specified; rather, the environment shapes them in such a way as to allow the organism to respond in a plastic manner to its changing demands. They specifically argue that this plasticity is driven by neural selection and that neural selection is implicated in the formation of most, if not all, of our interesting cognitive capacities (Buller 2005, 132-138; Buller and Hardcastle 2000, 315-316). One problem with this argument is that there are few reasons for thinking that neural selection is as ubiquitous as they believe it to be. In their defense, however, they probably do not *need* to rely on neural selection for the purpose of emphasizing the plasticity of the brain.

The mechanisms of neural *construction*, which will be discussed below and contrasted with neural selection, can also mediate brain plasticity. Hence, perhaps, their argument can be strengthened: they need not hold that neural selection, per se, is ubiquitous, but that *activity-dependent synapse formation* is ubiquitous, whether it be mediated by neural selection or neural construction.

The following is composed of six sections. After explicating the generalized selected effects theory and distinguishing it from other attempts to extend the selected effects theory of function outside of the context of evolutionary biology (Section 2), it will elaborate the claim that neural selection is a function-bestowing process in three main ways: *explication*, *defense*, and *prescription*. First, I will explicate the precise content of the claim that synapses can be *selected for*, by contrasting two different models of activity-dependent synapse formation: neural construction and neural selection (Section 3).

Secondly, I will defend the neural selectionist theory from a major criticism, namely, that it is vacuous (Section 4). Some critics of so-called "neural Darwinist" theories of the brain have argued that the concept of neural selection is not discriminating, in that, at some level of abstraction, all synapse formation can be seen as "selectionist" (Crick 1989, 247). This criticism rests on a failure to conceptualize appropriately what a selectionist view actually implies for modeling synapse formation (or the formation of other neural structures).

Thirdly, I will indicate the kinds of research projects that would be facilitated by considering various behavioral, cognitive, and neurobiological dispositions from the vantage point of the selected effects theory (Section 5). I will use drug addiction to suggest the empirical utility of the selected effects view and to gesture toward avenues for further research.

In conclusion, I will endorse a modest pluralism of function concepts and suggest that different concepts of function should be seen as complementary and the kinds of questions they suggest mutually illuminating (Section 6). Whereas the causal role theory of function emphasizes the need for mechanistic models of structures underlying complex cognitive abilities, the selected effects theory emphasizes the evolutionary and developmental production of such structures, and highlights the "reasons" for which they came into being. Moreover, understanding these etiological "reasons" may allow one to make predictions about the current causal roles of a neural structure, and to develop rational methods of biomedical or psychiatric treatment.

2 The generalized selected effects theory of function

A brief and very partial overview of the emergence of the selected effects theory of function will be given for the purpose of motivating the importance of the distinction between neural selection and neural construction (see Garson 2008). This will be followed by an explication of the generalized selected effects theory of function (GSE).

A Brief History of Functions

According to *etiological* theories of function, function statements are elliptical for causal-historical explanations for the origins of traits. These stand in contrast to *consequentialist* theories, in which the function of a trait is identified with a subset of its present-day capacities or dispositions. The main motivation for the etiological approach is the intuition that function statements (e.g., "biological trait *X* has the function *Y*") represent attempts to answer the question, "why is *X* there?" or "why does *X* exist?" For example, in telling my son that the function of the honeybee"s stinger is to protect the colony, I believe myself to be offering a causal explanation for why honeybees have stingers.[1] If protecting the colony did not explain why honeybees have stingers, I would be saying something false, or at least wildly misleading.

Of course, the claim that function statements are causal explanations encounters an immediate obstacle: how can the effect of a trait (protecting the colony) causally explain the trait itself (the stinger)? This is the problem of "backwards causation." A second major problem is the apparent *normativity* of function statements. By "normativity," I simply mean that something can possess a function that it is unable to perform (a trait can be "dysfunctional" or it can "fail to function"). A casual glance at any biological or biomedical journal reveals the extent to which the normativity of functions is presumed: pathologies, for example, are typically thought to result from a dysfunctional trait or a broken mechanism. Different hypotheses about the nature and source of these dysfunctions reflect different ways of modeling biological systems, and they suggest

different avenues for therapeutic treatment. Thus, the proper explication of "function" and "dysfunction" is far from trivial.

In 1970, the biologist Francisco Ayala succinctly formulated an answer to the problem of backwards causation (Ayala 1970). When an evolutionary biologist says that, "biological trait *X* has the function *Y*," he or she is citing an effect that *X* produced in the past, and that, as a result of natural selection, explains the present-day existence of *X*s. The function of the honeybee"s stinger is to protect the colony because, in the past, stingers protected colonies and as a result (in this case, by contributing to the inclusive fitness of honeybees with stingers) that trait was selected for *over* trait variants that did not have that consequence. As he put it, "the adaptations of organisms…are explained teleologically in that their existence is accounted for in terms of their contribution to the reproductive fitness of the organism" (Ibid., 9). Although Ayala does not draw attention to this fact, the etiological theory also explains the normativity of function: to say that a trait is "dysfunctional" or that it "fails to function" is simply to say that it is unable to perform the activity for which it was selected. Because the possession of a function is determined by history, and the disposition to perform that function by present-day structure and environment, it is perfectly evident how something can possess a function that it is constitutionally unable to perform.

Wimsatt (1972) and Wright (1973) adopted this solution in various ways. Neither Wimsatt nor Wright, however, included reference to natural selection in their respective *definitions* of the term "function." Wimsatt recognized that, "the operation of selection

processes is not only *not* special to biology, but appears to be at the core of teleology and purposeful activity wherever they occur" (Ibid., 13); however, he hesitated to build this observation into his definition of "function" as a necessary or sufficient condition (Ibid., 15-17). One reason was that he was able to come up with non-biological counterexamples, such as the possibility of "stellar evolution." If stars can undergo something like differential persistence – which seems to be a kind of "selection" process broadly construed – then such a theory would have the bizarre consequence that parts of stars have functions.

Wright (1973), similarly, did not make being selected for a necessary or sufficient condition for a trait"s having a function. Instead, he appealed to the very general concept of a "consequence etiology" in explicating "function" (e.g., Wright 1976, 116). Roughly, to say that a given object possesses a consequence-etiology is to say that one of the consequences it produced in the past figures into a complete account of its own continued persistence (either the persistence of the token or the type). Clearly, natural selection is one type of consequence etiology, but his definition is broad enough to include objects that do not undergo selection.

Wright"s failure to restrict functions to selected effects led to a series of apparently devastating objections, summarized famously in Boorse (1976). Many structures, Boorse showed, have consequence-etiologies without having functions. Imagine a hose in a laboratory that leaks a poisonous gas. Every time a scientist attempts to fix it, the scientist is knocked unconscious, which explains the persistence of the leak. In that case, Wright"s

theory has the consequence that the function of the leak is to knock out scientists. Such counterexamples can be multiplied endlessly. Obesity facilitates a sedentary lifestyle, which promotes obesity. Panic attacks promote a kind of hyper-vigilance to bodily sensations, which promotes further panic attacks. A stick floating in a stream that gets pinned to a rock may hold itself in place by the backwash it creates. Clearly, Wright"s formulation of "function" is overly general because it allows any self-perpetuating process to generate new functions.

In response to these sorts of objections, Neander (1983) and Millikan (1984) independently arrived at the position that the function of a trait consists (roughly) in the effect that it was selected for – precisely the place from which Ayala began. This formulation avoids Boorse-type counterexamples because none of them constitute selection processes: there is no obvious sense, for example, in which obesity is "selected for" *over* some other trait variant in an individual or population because it, rather than the variant, facilitated a sedentary lifestyle. Much of the functions literature in the 1990s was devoted to explicating, refining, or criticizing this basic insight in various ways (see citations in Section 1).

One pervasive assumption in the literature, an assumption shared by many of its critics, is that natural selection operating over an evolutionary time scale is the *only* natural process that can satisfy the selected effects theory (see citations in Section 1). Very few theorists attempted, in a rigorous and sustained manner, to use the selected effects theory to show how other kinds of natural processes of selection, such as antibody selection, neural

selection, or operant conditioning, might generate new functions. Moreover, the rare attempts that were made in this direction do not provide an adequate basis for identifying *neural* selection per se as a distinct function-bestowing process, or distinguishing it from other, non-selectionist, processes of synapse formation. I will briefly comment on the attempts made by Millikan (1984), Papineau (1987; 1993; 1995), and Godfrey-Smith (1992) in this direction to distinguish my own view from theirs.

Millikan notes explicitly that natural selection operating over an evolutionary time scale is not the only kind of function-bestowing process (Millikan 1984, 27). Certain learned behaviors, such as handshakes, can be seen as the result of a selection process. In her writing, operant conditioning appears to be the paradigm of this alternate kind of selection.[2] Here, the individual possesses a repertoire of behavioral patterns, some of which are differentially replicated (repeated or reinforced over others) because of their being correlated with a reward. The informal analogy between natural selection and operant conditioning can be traced back to Skinner himself (Skinner 1953, 430; 1981), and can also be construed formally (McDowell 2009). Whether and to what extent other forms of learning, such as classical conditioning, imitation, or enhancement learning, can be modeled as selection processes is an open question (e.g., Kingsbury 2008). Although Millikan does not apply the model to antibody selection, it would easily fit the model she envisions without modification, because antibody selection involves the differential replication of antibodies within the bloodstream over an ontogenetic time scale.

The main shortcoming with Millikan"s analysis is that it restricts functions to entities that undergo something like reproduction or "copying" (in her terms, to the members of "reproductively-established families"; Millikan 1984, 18). This restriction excludes neural structures from acquiring direct proper functions within the lifetime of the individual because neural structures do not reproduce or replicate (though they can be generated throughout the lifetime of some animals). Nonetheless, synapses, neurons, and entire groups of neurons undergo a process that is in every other respect analogous to selection in these paradigm cases (see Section 3).[3] Millikan did not seem to have considered the possibility that selection could act over individuals that are not capable of reproduction.

Millikan would not deny functions to unique, non-heritable neural structures. According to the view developed in her writings, such structures would not possess "direct" but "derived" proper functions. In short, a structure *X* possesses a derived proper function *Y* if it was produced by a mechanism that has the function *Y* and that typically performs *Y* through the production of structures such as *X* (Millikan 1984, 41-42; 1989; 288). The classic example is a novel pattern of camouflage on, say, a chameleon. Suppose (contrary to fact) that a given chameleon is placed in a novel environment and produces a novel pattern of camouflage – a token of a type that had never been exhibited in the history of the species – that allows it to conceal itself. This novel pattern, in its specificity, cannot be said to have a direct proper function because it was not selected for. Nonetheless, it was produced by a mechanism (the pigment rearrangers in the skin) that has the direct proper function of promoting camouflage and that typically performs this function by

producing structures such as coloration patterns. As a consequence, the novel coloration pattern would possess the derived proper function of promoting camouflage.

I accept the existence of such derived proper functions; in some sense, the two accounts complement each other rather than compete. However, it seems to me that my account of function (below), according to which *direct proper* functions can be bestowed by neural selection, is preferable for two reasons. First, my account of function is more parsimonious. It shows how a single process – namely, being selected for – can explain the emergence of functions in several different biological domains. Millikan"s analysis unnecessarily replaces reference to a single function-bestowing process with two different processes: *X* can be selected for, or, *X* can be produced by a mechanism that was selected for and that regularly carries out its direct proper function by producing entities such as *X*.

Secondly, Millikan"s account multiplies epistemic problems for the researcher who wishes to identify the etiological function of a trait. In some cases it would not be evident from which mechanism a neural structure derives its function. Consider the visual word form area (VWFA), located in the fusiform gyrus (see Garson 2011 for discussion). In English speakers it is differentially activated in the presence of words that conform to English spelling rules. It is not likely that this structure was selected for by natural selection because the written language has not existed for a very long time, and literacy does not seem to be associated with any fitness advantage. But neuroscientists tend to believe that this area does have a function, namely, the function of promoting the

recognition of the visual words of one"s own language. For Millikan"s analysis to be rendered consistent with neuroscientific usage, it would have to give the VWFA a *derived* proper function. But which mechanism is it that regularly fulfills *its* function by producing structures such as the VWFA, and what is the function of this mechanism? Arguably, the VWFA could be produced by a mechanism the function of which is to recognize visual shapes, or a mechanism the function of which is to facilitate communication, or simply a mechanism the function of which is to track correlations in the environment. It is not obvious how one would begin eliminating alternative hypotheses, or why one would want to assume such an explanatory burden in the first place.

Papineau (1987; 1993; 1995), like Millikan, attempted to broaden the selected effects theory beyond of the narrow ambit of natural selection. His main problem area was the domain of beliefs. In several works he described the possibility that beliefs themselves may undergo a kind of selection in which certain beliefs are "fixated" as a result of their consequences (Papineau 1987, 65-67; 1993, 44-48). For my purposes, however, I am concerned primarily with the way he applied this to neural structures themselves. He suggested that neural processes underlying belief formation may undergo a kind of selection that is sufficient for bestowing functions upon them: "I take it also that this neuronal mechanism was selected (reinforced, developed) *because* it produced that [disposition]" (Papineau 1995, 78). Godfrey-Smith (1992) endorsed a similar viewpoint, and argued that it is the *selective* character of learning which bestows teleological significance upon it: "It is important that the selective approach [to defining "function"] is

in no way tied to the genetic kind of biological evolution…A selective basis for functional characterization is available whenever learned characters are maintained within the cognitive system because of their consequences" (Ibid., 292).

Neither Papineau nor Godfrey-Smith (in these admittedly terse passages) develop this theory in a way that would distinguish neural selection per se from any self-perpetuating or self-amplifying neural process. This is because neither of them make a distinction between the claim that a synapse is "reinforced," "developed," or "maintained" because of its consequences and the claim that the synapse is *differentially* reinforced, that is, reinforced *over* some other synapse, because of its consequences. This would not only fail to identify neural selection as a distinct function-bestowing process, but it would render the theory almost completely vacuous in its application to the brain. This is because, as I will explain Section 3, *practically all activity-dependent synapse formation has a self-amplifying character*. That is, the frequent activation of a synapse typically leads to a structural change that raises the probability the synapse will be activated in the future. To attribute functions to *all* of these structures would essentially be Wright"s problem of self-perpetuating structures writ small. Thus, in order for the selected effects theory to be applied in a non-vacuous manner to synapse formation (and the formation of other neural structures), a more specific theory is required.

A Generalized Theory of Function

The basic shortcomings in these attempts to apply the selected effects theory outside of the evolutionary context can be resolved in a fairly simple way, through what can be called the generalized selected effects theory (GSE):

> (GSE): The function of a trait consists in that activity which historically contributed to its differential reproduction or its differential retention within a biological population.

This characterization of function depends on the notion of a biological population. Moreover, the way I distinguish neural selection and neural construction also relies on the notion of a biological population. Hence, a few words on the subject are warranted. As Millstein (2009) notes, few philosophers of biology have addressed extensively the issue of what a biological "population" is, despite the importance of the term in biology. One way of motivating the importance of the explication of "population" is that whether something is or is not a process of natural selection depends upon how populations are individuated, since selection requires variation within a population. The reason that the peppered moths famously studied by Kettlewell can be said to have undergone natural selection is because the light- and dark-colored variants were part of the same "population." If one stipulates that the light- and dark-colored moths each constitute a separate "population," then, relative to that stipulation, natural selection could not be said to have taken place (ibid., 268).

As Millstein suggests, there are three main ways of characterizing biological populations: via spatial boundaries, causal interactions, or history. (All of them assume that all of the members of a population belong to the same species.) The first is by reference to spatial (and perhaps temporal) boundaries. According to this construal, two strains of bacteria in a test tube that do not interact form a "population" by virtue of being within the same spatial enclosure. According to the second construal, two individuals are members of the same population because of their typical patterns of interaction. But which sorts of interactions are relevant to defining populations? The most obvious candidates are interactions that affect fitness, both competitive and cooperative. According to the third construal, populations are identified by their shared history, i.e., by the fact that sets of individuals occupy the same position on a genealogical tree. There is no deep discrepancy between the second and third approaches, as having a shared genealogy is one outcome of interactions between individuals (ibid., 270).

In the following, I accept (as does Millstein) an interaction-based definition of "population," in which fitness-affecting interactions are the relevant sort of interaction. This approach will be used as a working characterization of "population" and not as a rigorous definition equipped with necessary and sufficient conditions. My main concern is to have a starting point to get GSE off the ground, and to provide an adequate distinction between neural selection and neural construction. Different explications of what constitutes a "fitness-affecting interaction" could lead to different ways of identifying the members of a population; it is not my intention to develop the notion in a more precise way. The strategy of defining "population" in terms of fitness-affecting

interactions is similar to D.S. Wilson"s strategy of defining "group" (or "trait-group"), where a trait group includes all of the individuals the fitnesses of which are affected by a given trait, regardless of the spatial distribution of members of the group (Wilson 1975; also see Sober and Wilson 1999, 92-98).

There are two benefits of appealing to this interaction-based approach to population in defining "function." The first is that it avoids counterintuitive function ascriptions that would otherwise result from the bald appeal to differential reproduction or retention, as will be described below. The second is that it provides a distinguishing hallmark between neural selection and neural construction. An essential feature of neural selection is that there are fitness-affecting interactions between synapses, neurons, or groups of neurons, where "fitness" in this context refers to the propensity of the synapse, neuron, or group of neurons to persist (or "survive"). Different neural structures can affect the fitnesses of other neural structures, thereby constituting populations. Neural construction, however, does not necessarily involve such fitness-affecting interactions. I will come back to this point in Section 3.

One cost to this particular definition of "population" is that it would not count certain processes as "selection processes" even though they would normally be counted as such. In Lewontin"s (1970, 1) famous example, if there are two strains of bacteria in a test tube and unlimited resources, and one strain has a faster division time than the other such that it is increasing in relative frequency in the tube, then natural selection is taking place. However, supposing that there are unlimited resources available to both strains of

bacteria, and there are no fitness-relevant interactions between the two different strains, the interaction-based definition would not count both strains as part of a single "population" despite their spatial proximity.[4] From an evolutionary point of view, the two strains may as well occupy separate tubes. As Sober and Wilson (1999) put the point in reference to groups, "Individuals belong to the same group because of their interactions, not because they are elbow-to-elbow (93)."

There are three main differences between GSE and other, closely related, etiological views. First, in contrast to Wright (1973), and similar to Millikan (1984) and Neander (1983), it insists that being selected for is necessary for having a function.[5] The expression "contributed to its differential reproduction…or its differential retention" is intended to reflect the fact that the activity must have done something to promote its being selected for *over* some other trait.[6] Second, in contrast to Millikan"s formulation, it does not restrict selection to entities that undergo something like "reproduction" or "copying" (that is, differential "retention" also suffices for having a function).

Third, it restricts functions to members of biological populations. This allows it to avoid the kinds of counterexamples developed by Wimsatt (1972, noted above), Schaffner (1993, 383), or Kingsbury (2008, 496), in which various systems undergo something like differential retention but would not intuitively qualify as possessing functions. Kingsbury (ibid.) imagines a collection of large rocks on a beach, where each rock varies in hardness and therefore in erosion resistance. This collection undergoes differential persistence, yet none of its members have functions. This is not a problem for GSE, for two reasons.

First, GSE is a theory of *biological* function and thus it is appropriate to restrict it to biological systems (Garson 2011, 558). Secondly, and more importantly, a collection of rocks is not a *population* in the sense characterized above. In order for the collection of rocks to constitute a population there would have to be fitness-relevant interactions between the individual rocks that make it up. The fitness of a given rock – here construed in terms of some measure of persistence – would have to make a positive or negative contribution to the fitnesses of other rocks. But it doesn"t, since the fate of each rock is independent of the fate of the others. The same point could be made about Wimsatt"s example of stellar evolution, or the different ball bearings in Schaffner"s imaginary "cloner." They only count as "populations" according to some measure of spatial proximity rather than by virtue of fitness-relevant interactions.

3 Neural selection and neural construction

If selection processes are operative at the neurobiological level, there exists warrant for assigning etiological functions at that level – that is, to structures that may be unique, non-heritable, and capable of rapid reorganization in response to novel environmental demands. This section will begin by considering neural selection operating at the level of the *synapse* ("synapse selection") as this is the paradigm case of neural selection and, in the context of synapse formation, there is a clear antithesis, namely, neural construction. It will then describe selection processes at higher levels, and will contrast neural selection and neural construction. The presentation of both views can help to illuminate the content of the former by providing a well-defined contrast.

Activity-Dependent and Activity-Independent Synapse Formation

There are two kinds of mechanisms that give rise to synapses: "activity-independent" and "activity-dependent." To say that a neuron"s pattern of connectivity – that is, its pattern of divergence (the set of neurons it innervates) and its pattern of convergence (the set of neurons that innervate it) – is "activity-independent" is to say that this pattern is not based on the activation of that neuron, i.e., the production of electrical potentials and the release of neurotransmitter (or other signaling molecules in the case of gap junctions). To say that the pattern is "activity-dependent" is to say that the pattern is based, in part, on that neuron"s being activated.

An example of a theory according to which synapse formation is largely activity-independent is the chemoaffinity hypothesis associated with Roger Sperry. According to this view, each neuron possesses a specific chemical "marker," established genetically or in the early stages of neural development, and that neuron is guided to a specific target that bears an identical or complementary such marker (e.g., Sperry 1951; 1963; also see Meyer 1998). A second type of theory that holds synapse formation to be activity-independent represents a variation on the strict chemoaffinity hypothesis. It holds that this chemical signal or "marker" is not unique to a given neuron but to a class of neurons, and that the affinity between an innervating neuron and its target exhibits gradation in strength (Meyer and Sperry 1976). A third type of theory that characterizes synapse formation as activity-independent refers to the role of purely mechanical constraints, such

as substrate guidance, for explaining the initial pattern of connections (e.g., Scholes 1979). Such constraints are particularly important in early neural development.

In the neurobiological context, "learning" is often used very broadly to refer to activity-*dependent* changes in synaptic structure (e.g., Kandel et al. 2000, chapter 63). Learning, in this sense, involves a relation of dependence between neural activity and synapse structure: activity partly determines structure. But how, precisely, does neural activity determine synapse structure? Two very general views that attempt to answer this question are "neural selection" and "neural construction." To avoid confusion, however, it is crucial to note that the particular mechanism that is typically referred to as "neural selection" in the literature refers to one *paradigm type* of selection process, namely, that in which there are competitive interactions between different synapses, neurons, or groups. Strictly speaking, competitive interactions are not necessary for a process to qualify as a neural selection process. All that is required is the differential retention of synapses, neurons, or groups that belong to the same population. In the following, I will use the term "neural selection" as it is used in the literature, namely, to refer to the paradigm case, which involves competitive interactions. In Section 4, I will provide an example of a neural selection process that is not mediated by competitive interactions.

Synapse Selection

Neural selectionists typically view synapse formation as a two-stage, iterated process (Changeux and Danchin 1976; Changeux 1985; Edelman 1978; 1987; Gazzaniga 1992;

Edelman and Tononi 2001). The first stage corresponds to the activity-independent formation of new synapses. This produces an initial pattern of connectivity that is, to a large extent, both *random* and *exuberant*. That is, this process creates a large repertoire of synaptic variation, much of it non-adaptive. The second stage corresponds to the reduction of variation via the "competitive" elimination of certain synapses (in a sense that will be clarified below). This latter stage is an activity-dependent process. A representative quote nicely summarizes this perspective: "To learn is to stabilize preestablished synaptic combinations, and to eliminate the surplus" (Changeux 1985, 249). Crucially, this two-stage process does not occur only once in the development of the individual. It is an iterated process, with cycles of proliferation and elimination in different brain regions at various stages of the individual"s life.

Note that, strictly speaking, synapse selection does not *require* that the initial set of synapses is in some sense "randomly produced," that is, that it involves activity-independent proliferation of new synapses.[7] It is certainly possible that activity-*dependent* proliferation of synapses could be followed by the competitive elimination of some of them. However, as will be noted below, activity-dependent synapse formation has a "directed" quality that reduces the need for a subsequent selection process. In other words, in neural construction, a synapse is strengthened as a result of being frequently activated, or a new synapse forms as a result of the frequent co-activation of adjacent neurons. Consequently, the probability that a synapse will be frequently utilized, given that it was produced by an activity-dependent process, is greater than the probability that a synapse will be frequently utilized, given that it was produced by an activity-

independent process. This "directed" quality of neural construction minimizes, though does not entirely obviate, the need for the subsequent elimination of infrequently used synapses.

The concept of "neural selection" itself, as an explicit analogy to natural selection, was first proposed and developed in two foundational papers. The clearest early expression of the view is Changeux and Danchin (1976), though the ideas were expressed in an earlier paper (Changeux et al. 1973), the accessibility of which was dampened by a cumbersome formalism. This view emphasized neural selection in terms of competitive interactions at the level of the synapse. Changeux and Danchin"s work was prompted by the earlier work of English neurophysiologist J. Z. Young on learning in cephalopods.[8] In his book, *A Model of the Brain*, Young proposed a simple model of learning according to which various synapses initially elicit various behaviors in a random fashion; those behaviors in the organism"s repertoire that are "successful" produce a signal that leads to the differential reinforcement of the synapse controlling that behavior and the differential weakening or elimination of the synapses that control contrary behaviors (Young 1964, 285). Young"s model can be seen as an alternative to the Hebbian model which describes only the activity-dependent amplification of existing synapses or the formation of new synapses between simultaneously active neurons (see below, where I develop the contrast between Young"s selectionist model and Hebb"s constructionist model of synapse formation).

Two paradigm cases that crop up frequently in the literature on synapse selection are experience-dependent plasticity in ocular dominance in visual cortical neurons, and the formation of the neuromuscular junction. In the following, I will discuss ocular dominance (see Garson 2011 for discussion of the neuromuscular junction and references). Neural selection has also been implicated in the formation of brain regions underlying filial imprinting and the olfactory system, amongst others (see Lichtman et al. 1999 for a review).

A competitive interaction model of ocular dominance plasticity was initially suggested by experiments carried out by David Hubel and Torsten Wiesel in the 1960s (Wiesel and Hubel 1963; Hubel and Wiesel 1965). These experiments consisted in depriving newborn kittens of visual stimulation in one eye (monocular deprivation) for the first few months of life, and recording electrical activity from single cells in the visual cortex (Wiesel and Hubel 1963). In normal cats, most of the cells of the visual cortex are responsive to visual stimulation from either eye; specifically, about 80% of those cells are "binocularly driven," although a small proportion are exclusively responsive to stimulation of one eye or the other. This is referred to as the "ocular dominance profile" of the cell. In monocularly-deprived cats, the vast majority of cells respond exclusively to stimulation from the non-deprived eye alone; they are "monocularly-driven." This results from a plastic reorganization of the visual cortex that has the effect of maximizing the visual acuity of the non-deprived eye.

The ocular dominance profile of one cell is not completely independent of its neighbor, but cells with the same profile tend to cluster together in groups called "ocular dominance columns" (or "bands"). In the normal visual system these groups can be made to appear as a pattern of "stripes" along the visual cortex of about equal width. The results of monocular occlusion can be visualized in terms of the unequal widths of the ocular dominance columns associated with either eye (Hubel and Wiesel 1972; Hubel et al. 1977).

But how does this relate to the notion that there exists an active "competition" between the neurons that carry information from either eye? Unlike kittens that have undergone monocular deprivation, kittens that have been exclusively dark-reared for the first several months of life retain largely the same degree of binocularity as normal kittens. This implies that the results of monocular deprivation cannot be explained on the assumption that connections from the deprived eye degenerate as a function of disuse. Rather, as Wiesel and Hubel put it, there is a "competition" between the neural connections from the deprived eye and the non-deprived eye (Ibid., 1015); that is, the loss of connections from one eye somehow results from the activation of the other eye. Specifically, there seems to be a competition between geniculocortical synapses (synapses between neurons in the lateral geniculate nucleus and neurons in the visual cortex).

Various visualization techniques support this "competitive" paradigm. Rakic (1976) used a staining method to demonstrate that, in the fetal Rhesus monkey brain, geniculocortical axons are diffusely distributed and intermixed in the visual cortex, and that the

segregation of these axons into ocular dominance columns begins in the second half of gestation. This is consistent with a selectionist model in which synapses are initially distributed in a random and diffuse manner, and the mature synaptic structure results from the progressive withdrawal of some geniculocortical axons and strengthening of others, likely as a result of spontaneous activity of the lateral geniculate nucleus (see Katz and Shatz 1996 for discussion). This research was also important because it suggested that competitive interactions mediate not only cortical plasticity in response to monocular occlusion, but the normal development of ocular dominance columns as well, though this remains contentious (Price et al. 2011, 210).

The existence of this "eliminative" process – specifically, the retraction of geniculocortical axons associated with the deprived eye – can be visualized at the level of the single neuron as well. As Antonini and Stryker (1993a) show, upon subjecting newborn kittens to monocular deprivation, the retraction and elimination of geniculocortical projections is initiated very rapidly; within 6-7 days after monocular deprivation, geniculocortical axons controlled by the deprived eye are shorter in length and have fewer branches than those controlled by the non-deprived eye. This is consistent with a selectionist model according to which the unused synapses selectively "withdraw" from the target.

Many computational models utilize competitive interactions between synapses for modeling experience-dependent cortical plasticity following damage to afferent neurons (Buonomano and Merzenich 1998, 175-179). One rule utilized in such models is

postsynaptic normalization, which requires the total strength of all inputs to a target neuron to remain constant. This has the consequence that the strengthening of one synapse entails the weakening of others, and vice versa (van Ooyen 2011, 321-322); thus, the use of postsynaptic normalization is a way of incorporating competitive interactions into computational modeling. However, the use of postsynaptic normalization in computational models does not answer the question of the biological mechanisms that might give rise to it. *If* postsynaptic normalization often mediates cortical plasticity, and *if* competitive interactions between synapses provide the mechanism for postsynaptic normalization, then synaptic selection would prove to be a very common mechanism of cortical plasticity.

A major question that arises here is the following: if synapse selection involves a "competitive" process, then what precisely is the "resource" over which the synapses are competing? One possibility is that synapses compete for a neurotrophic factor (NT) synthesized by the target neuron and required for the maintenance of the synapse (Harris et al. 1997; 2000; Elliott and Shadbolt 1998; 2002). Two pieces of evidence involving the artificial infusion and deprivation of NT support the hypothesis that ocular dominance formation involves a "competition" for NT (Harris et al. 2000). The infusion of NT prevents the formation of ocular dominance columns; this suggests that the unlimited availability of NT obviates the need for competitive interactions. Correspondingly, the deprivation of NT is associated with the decay of inputs from both eyes, which suggests that NT is required for the maintenance of normal synapses. Another competitive hypothesis might involve direct negative interactions between innervating neurons (ibid.;

also see Price et al. 2011, 292-293, and below, on the use of such models for explaining competitive cell death).

Selection at Higher Levels

Just as there are levels of selection in the evolutionary context (e.g., genic, chromosomal, individual, group, and species), there are levels of selection in the neural context: theoretically, selection is possible at the level of the synapse, at the level of the entire neuron, and at the level of groups of neurons. Each of these categories will be summarized below.

As noted above, the explicit use of "selectionist" models was first developed by Changeux and colleagues to describe selection at the level of the synapse (Changeux et al. 1973; Changeux and Danchin 1976). To my knowledge, no neuroscientist has explicitly claimed that synaptic selection is speculative, as there are a host of well-documented examples (see Lichtman et al. 1999 for a review), though its prevalence is questioned (see below). As a consequence, it has earned its place in most graduate-level textbooks as an important mechanism of synapse formation.

A "higher" level of selection takes places at the level of individual neurons. Selection of neurons is primarily exemplified by the phenomenon of neural cell death or apoptosis, which refers to a phase of neural proliferation and migration followed by a period of widespread cell death during embryonic development in vertebrates (Cowan 1973; 1978;

Oppenheim 1991; Johnson and Deckworth 1993; Pettmann and Henderson 1998; Deppmann et al. 2008). Hamburger and Levi-Montalcini (1949) first identified the generality of this phenomenon by observing motor neuron degeneration in the spinal cord of chicks.

The primary function of neural cell death appears to be a quantitative one: it matches the size of a given group of neurons with the size of its innervation field, that is, the population of target neurons or receptors that the group innervates (Cowan 1973). This is suggested by the long-attested fact that increasing the size of the innervation field through limb transplantation increases the number of motor neurons that survive cell death, and decreasing this field through limb extirpation decreases it (Detwiler 1936; also see Hollyday and Hamburger 1976). In addition to this quantitative function, it may also serve to eliminate some connections that have been formed by "developmental errors." For example, Clarke and Cowan (1975; 1976) injected a retrograde tracer into the eye of the chick embryo and showed that a small number of neurons in the ipsilateral, rather than contralateral, isthmo-optical nucleus were labeled with the tracer (these would represent such "errors"). About 80-90% of those labeled neurons die in early development, suggesting that neural cell death performs the qualitative function of eliminating neurons that innervate the "wrong" eye (Clarke and Cowan 1976, 144).

The quantitative function of neural cell death may be mediated by a "competitive" mechanism. The simplest theory is that neurons that successfully innervate a target are preserved, and those that fail to innervate die (Hamburger 1958, 399). Retrograde

labeling techniques, however, have shown that neurons that successfully innervate a target are also subject to cell death (Clarke and Cowan 1975; 1976). This suggests the possibility that, like synapse selection, a competitive process between neurons that involves the uptake of a diffusible trophic substance mediates cell death (Cowan 1973; 1978, 166). For example, nerve growth factor (NGF), a protein originally isolated from snake venom, was found to contribute to the survival of sympathetic and sensory ganglia in vitro (Cohen and Levi-Montalcini 1956; Levi-Montalcini and Cohen 1960), and was later found to occur naturally in rat sympathetic target tissue (Ebendal et al. 1980). Since that time, many different NTs have been identified (Walicke 1989; Huang and Reichardt 2001). NTs are currently believed to sponsor the survival of neurons not by enhancing cell metabolism, but by suppressing a set of genes that are responsible for self-destruction (Yuan and Horvitz 1990).

To the extent that neural cell death results from competition, it is not known precisely what variable feature of a neuron gives it a "selective advantage," that is, what trait confers differential survival onto that neuron. This is determined by what, precisely, the "limiting resource" turns out to be. On the one hand, according to what might be termed the "production hypothesis" (Oppenheim 1989, 253), NTs such as NGF are *synthesized* in limiting quantities. Hence, any mechanism that promotes the differential uptake and retrograde transport of NTs would be "selected for," in that it would not only increase intracellular availability of those NTs, but also deplete the target source and thereby deprive other neurons of trophic support (Davies et al. 1987, 358; Bothwell 1995, 245). On the other hand, according to what might be called the "access hypothesis" (Oppenheim

1989, 254), it is not limited synthesis per se which drives competition but a limited number of available synaptic sites on the target. According to this hypothesis, neural cell death would result from the "competition" for space. A third model has recently been proposed which involves direct negative interactions between neurons. According to this view, neurons with high trophic signaling kill neurons with low trophic signaling through the release of an apoptosis signal (Deppmann et al. 2008). Again, however, the basic theory of neural cell death, and the mediation of this phenomenon through some type of "competition," is not particularly controversial.

A higher level of neural selection is "neural group selection." This is the theory associated with the work of Edelman (e.g., 1987; Edelman and Tononi 2001) and which has been subject to the most severe criticism of the three (e.g., Barlow 1988; Purves 1988; Crick 1989). This theory holds that basic cognitive capacities such as pattern recognition result from selection of groups of neurons. According to this view, one outcome of genetic and epigenetic processes is the construction of large repertoires ("primary repertoires") of neural groups, each group consisting of 50 to 10,000 neurons. Each group in the repertoire exhibits a different internal pattern of connectivity but responds in various degrees to the same stimulus pattern (hence they exhibit "degeneracy"). All groups in the repertoire are "isofunctional" because they share a similar response profile, but they are "nonisomorphic" because they differ structurally (Edelman 1978, 64-65). Selection acts over this primary repertoire in the following way: the neural group that responds most specifically to the stimulus pattern that defines the repertoire is differentially

strengthened (that is, its intraspecific pattern of connections is strengthened over that of the other groups).

Edelman first wrote about the concept of neural selection in 1978 (Edelman 1978) and explicitly drew upon Changeux and Danchin"s work. Edelman"s interest in selectionist ideas, however, had a different origin than Changeux"s. Edelman"s scientific career began in immunology, and he was a supporter of the "selectionist" approach to antibody production that was proposed by Jerne (1955) and developed in several papers and books (Lederberg 1959; Burnet 1959; Jerne 1967). In brief, the "clonal theory of antibody production" holds that a mechanism of genetic recombination is responsible for the random production of a large variety of antibody molecules that circulate at low levels of the bloodstream. When the antibody comes into contact with the antigen specific to it (sharing the same conformation), that antibody is differentially replicated throughout the bloodstream. Jerne (1967) proposed this "selectionist" theory as an alternative to so-called "instructionist" views according to which the antigen somehow impresses its form on a non-specific or "plastic" antibody and induces the latter to adopt the correct conformation pattern. Although Edelman"s previous work hinted at the possibility of an analogy between the selectionist approach to immunology and selection in the nervous system (e.g., Edelman 1967, 199; 1975) he did not seem to have developed the analogy in any detail prior to the 1978 paper.

To my knowledge there are no well-documented cases of neural group selection. Even Edelman, in defending his view, cites evidence for synapse selection to support his theory

(see Edelman and Tononi 2001, 84). Despite the lack of evidence for neural group selection, it would be a conceptual error to disregard neural selectionist views on the whole because of shortcomings of Edelman"s own view. The error would not be unlike rejecting the theory of natural selection on the basis of shortcomings with group selection.

Neural Construction

The second main position on how neural activity translates into synapse structure is called neural construction. Proponents of neural construction hold that the formation of new synapses is itself an activity-dependent, non-random process (e.g., Purves 1994; Purves et al. 1996; Quartz and Sejnowski 1997). Hence, neural construction emphasizes the role that activity plays in the formation and strengthening of new synapses, rather than the elimination of existing ones. For example, suppose that neuron *A* synapses onto neuron *B*. The activation of *B* by *A* may trigger the growth and extension of new dendrites on *B* and new axon terminals on *A*, or the upregulation of membrane channels on *B*. This would increase the strength of the connection between *A* and *B* in an activity-dependent manner. Moreover, this need not be a selection process, as it typically involves merely the multiplication of new synapses, or the strengthening of existing ones, rather than the *differential* retention of synapses, where all of the synapses belong to the same population.

These newly formed neural projections may also branch, extend, and form synapses with neighboring neurons. In this case, the joint activity of *A* and *B* can promote the formation of new synapses without selection. In sum, constructionists view neural growth in terms of the gradual, progressive, and activity-dependent elaboration of novel synaptic structures and circuitry on an "as-needed" basis, rather than the elimination of randomly-formed "excess" circuitry. Purves (1994) gives a clear statement of this viewpoint: "…activity-dependent growth provides a richer and more consistent framework for thinking about neural development than the now popular idea that we start life with an initial excess of connections and then select from this surfeit by competitive mechanisms akin to natural selection. Rather, the brain builds the circuitry it needs during its progress to maturity" (Ibid., 93-4).

Constructionists do not deny completely the selective elimination of existing synapses. They acknowledge that the activity-dependent formation of new synapses is partly stochastic and "error-prone," and hence may require the selective elimination of useless or maladaptive ones (Purves 1994, 68; Quartz and Sejnowski 1997, 550). The "directed" quality of synapse formation, however, minimizes the need for a consequent phase of selection. In other words, neural construction typically ensures that synapses are not strengthened or formed unless they are likely to be frequently activated.

Hebb (1949, esp. Chapter 4) can be read, in retrospect, as a paradigm statement of neural construction. "Hebb"s law," as it came to be known, simply states that when one cell, *A*, is repeatedly involved in the firing of another, *B*, *A* and *B* undergo some structural change

such that the probability of *A*"s causing the firing of *B* is increased (Hebb 1949, 62). Although he remained somewhat neutral about the mechanism of this change, he proposed to account for it by supposing that, as a result of *A*"s causing the firing of *B*, *A*"s axon terminals grow to cover a greater area of the dendrites or soma of *B* (ibid.). In large groups of adjacent neurons, a pattern of synchronous firing amongst various neurons in the group would lead to the progressive growth or strengthening of diverse connections between them. This model does not logically require the existence of "unused" synapses that are weakened or eliminated as a result of the strengthening of others. The mechanistic model that Hebb proposed to explain "Hebb"s law" is not a selectionist model.

One might think that, according to this model, there could be "differential proliferation" of synapses (*without* differential elimination), and thus that neural construction, strictly speaking, can also constitute a synapse selection process.[9] This conception, if true, would substantially blur the traditional boundary between selection and construction. But the most obvious example of the "differential proliferation" of synapses would involve selection at the level of the *neuron* and not at the level of the synapse, which is what is at issue here. This is because the only entity the differential "fitness" of which is enhanced by the mere multiplication of new synapses is the neuron itself (where "fitness" is construed as the disposition to persist). This point can be illustrated through the following example.

Suppose there are two neurons, G (for "gregarious") and R (for "reserved"). G and R vary in that G is more disposed to form new synapses on an "as-needed" basis than R. Suppose, moreover, that there are fitness-relevant interactions between G and R (or between their constituent parts) such that they belong to the same population. Over time, G forms more synapses than R. Note, however, that the mere fact that G is disposed to form more synapses than R does not affect the fitness of any of G"s existing synapses, where fitness is measured in terms of the disposition to persist. (By analogy, the mere fact that someone is disposed to make a lot of friends does not affect the duration of any particular one of those friendships.) Hence, this would not qualify as a selection process at the level of the synapse.

This scenario would be more reasonably characterized as involving the differential fitness of neurons. That is, it may be that G itself – the entire neuron – has a higher chance of avoiding cell death than R as a result of its many synapses. Empirically, the ability to innervate a target and utilize NT lowers the probability that the neuron will undergo cell death. If so, then this model describes selection at the level of the individual neuron, and not at the level of the synapse. Nobody has ever doubted that neural construction at the level of the synapse can set the stage for neural selection at higher levels, i.e., the level of the neuron or the level of the group.

One might also construe this scenario as involving selection at the level of the axon, that is, in terms of a selection process that takes place between G"s axon and R"s axon. But in this example, R"s *axon*, per se, is not jeopardized by the fact that it makes fewer synaptic

contacts than G"s axon. (Of course, if the entire neuron R dies because of the lack of synapses, then R"s axon dies. But this is a case of selection at the level of the neuron and not at the level of the axon.) Therefore, this "axonal selection" model requires a different construal of the notion of "fitness." Here, the "fitness" of an axon would not consist in the probability of survival of that axon, but in the number of branches that it has. In this respect it would be similar conceptually to one construal of "clade selection," which would measure the "fitness" of a clade by the number of branching events it sponsors (Okasha 2003, 747; though Okasha does not actually hold this view). Perhaps this can be seen as an intermediate level of selection, falling between selection at the level of the synapse and at the level of the neuron. Though I am not averse to developing the theory in this direction, the only point I wish to make is that axonal selection would seem to require a fairly unconventional interpretation of "fitness." Moreover, it would still be selection acting at a different level than that of the synapse. Hence, the differential multiplication of new synapses wrought by neural construction should not be thought of as a kind of selection process that takes place at the level of the synapse.

One of the central pieces of evidence that some neural constructionists have raised against neural selection is that the total number of synapses, axonal branches, and dendritic branches in the human brain *increase*, rather than *decrease*, from birth to puberty. If the development of the mature synaptic structure of the brain merely involved the proliferation of synapses early in development, followed by a long phase of competitive elimination, one would expect total synapse number to decrease, not to increase (Purves et al. 1996, 461). This argument misconstrues the implications of neural

selection models. It presupposes that the two stages of neural selection – proliferation of synapses followed by elimination – occurs only once in development. Advocates for neural selectionist views have emphasized that the two stages envisioned in synapse selection are iterative (Changeux 1997; Sporns 1997a). Synaptogenesis is not arrested after the first stage. The further branching and growth of new projections, after an initial round of competitive elimination, represents a re-iteration of the first stage of activity-independent, random, and exuberant growth. Synapse selection via competitive interactions is compatible with the total increase in synapse number from birth to puberty.

A Conciliatory View

There exist well-documented cases of both processes, selection and construction, in the brain, and to this extent there is no genuine dispute regarding the bare existence of either mechanism.[10] In many cases, both processes operate concurrently. The debate involves the relative significance of either mechanism, much like the adaptationism debate in evolutionary biology. Some proponents of selection, such as Changeux (1985; 1997), Edelman (1978; 1987), Gazzaniga (1992), and Sporns (1997a; 1997b), argue that virtually all synapse formation is selectionist in character, while those such as Purves (1994), Purves et al. (1996), and Quartz and Sejnowski (1997), emphasize its constructive character. Katz and Shatz (1996), LeDoux (2002), Black and Greenough (1986; 1997), and Elliott and Shadbolt (1997) emphasize the concurrent operation of both processes and hold that they are complementary or even inextricable from one another.

The work of Antonini and Stryker (noted above) provides a clear model of the complementarity of both processes. After subjecting newborn kittens to monocular deprivation, Antonini and Stryker showed visually that geniculocortical axons associated with the deprived eye retract very rapidly, which supports a selectionist view. In contrast, the geniculocortical axons controlled by the non-deprived eye are longer and have more branches than those of a normal kitten (1993a, 1820). This suggests that ocular dominance plasticity is not merely controlled by an eliminative process, but by a "constructive" process as well, in which the activity-dependent branching and growth of new axon terminals occurs on an "as needed" basis. On the basis of their work, Antonini and Stryker (1993b) embrace the conciliatory view that "normal development [of ocular dominance columns] thus appears to involve both selective elimination of widely extended branches and considerable growth and elaboration" (Ibid., 3572).

One implication of GSE is that *neural selection produces novel brain functions, but neural construction does not*. At best, neural construction only serves to amplify existing functions. One consequence is that if neural selection is very frequent, then novel brain functions will frequently arise. If it is infrequent, then novel brain functions will arise less frequently (though there are other processes that give rise to novel brain functions, namely, natural selection operating over an evolutionary time scale). Unfortunately, nobody knows, relatively speaking, how ubiquitous neural selection is as opposed to neural construction. There are well-documented examples of each, and I don"t know of any compelling theoretical arguments to the effect that one or the other is vastly more prevalent, despite the insistence to the contrary by some of the advocates on either side.

4 Is 'Neural Selection' Vacuous?

One particular criticism that has been leveled against "neural selection" is that it is vacuous, or non-discriminating. This criticism was leveled by Francis Crick, who stated summarily that, "almost everybody"s theory could be called a theory of synaptic selection" (Crick 1989, 247). Crick"s argument for this claim was that any explanation of the mature synaptic structure of the brain must show how it comes about that some synapses are formed and strengthened while others are weakened or eliminated, which is a kind of "selection." But this would render the theory vacuous. More problematic for the theory of function proposed here would be the implication that almost every activity that the brain facilitates via the coordinated action of a group of neurons would constitute a "function" of that group, including paradigmatically dysfunctional states such as *grand mal* seizures. It would be, as noted above, a repetition of Wright"s problem of self-perpetuating structures.

This criticism stems from a failure to appreciate what "selection" means in the neurobiological context. Selection takes place over a biological population. That means that in order to figure out whether or not a certain neural process is a "selection" process, one must specify which level of selection is being referred to (synapses, neurons, or groups of neurons) and then specify how the "populations" are identified. According to the approach described above (Section 2), a collection of individuals constitutes a "population" if the individuals engage in fitness-relevant interactions. For neural

structures, this involves interactions that affect the disposition of that neural structure to persist.

Though these interactions can be competitive or cooperative, one very important type of fitness-relevant interaction between neurons is a competitive one. As noted above, this particular form is so prominent that the term "neural selection" is sometimes used synonymously with "neural competition," though neural competition is a subset of neural selection. It is possible for a collection of synapses to form a population by virtue of their cooperative interactions. As Price et al. (2011, 275) note, most cells in the visual cortex are binocularly driven. That means that they accept synapses from different axons that carry information from different eyes. In some cases these diverse synapses can have a mutually strengthening effect. This is called "associative plasticity." Moreover, if selection does occur amongst the members of this population, it need not take the form of a competition over a limiting resource. For example, suppose neurons A and B synapse onto the same neuron, C, and that these synapses have had a mutually reinforcing effect. Suppose, moreover, that the resource R provided by C is for all practical purposes unlimited. Suppose, finally, there is variation such that the A-C synapse is able to utilize R more effectively than the B-C synapse. For example, one may imagine that the A-C synapse is more efficient at converting R into new axon terminals that lead to the strengthening of the A-C connection, but in such a manner that does not directly jeopardize the B-C synapse. This would count as differential retention amongst synapses that belong to the same population without involving competition.

The fact that fitness-relevant interactions are necessary for neural selection can be used to distinguish between neural selection and neural construction. It is easy to show this by envisioning a case in which there is a collection of synapses that do not have fitness-relevant interactions, that is, that do not constitute a population. Suppose there are four neurons, A, B, C, and D. A synapses onto C, and B synapses onto D. As a result of the frequent co-activation of A and C, the synapse between them is strengthened (neural construction). As a result of the non-coordinated activation of B and D, the synapse between them is weakened. As long as there are no fitness-relevant interactions between A and C, then this is not a selection process. Even if A and B are adjacent to one another, they may as well be on other sides of the brain or even within the brains of different people, because as noted in Section 2, *physical proximity between a collection of individuals does not make that collection into a biological population.* It is simply not true that, "almost everybody"s theory could be called a theory of synaptic selection" (Crick 1989, 247).

5 Utility of neural selectionist views in the context of neuroscience

There are a few reasons why the consideration of selected effects functions in the context of neuroscience is important for research, prediction, and biomedical or psychiatric treatment. In some cases, understanding the "reasons" (*qua* selected effects) for a certain belief, behavioral disposition, or neural structure is crucial for making accurate predictions about other neural or cognitive processes that take place within the individual. In the biomedical and psychiatric contexts, this is important for constructing rational

methods of treatment. An interesting example can be taken from the science of drug addition.

Drug addiction is facilitated by the coordinated actions of many different brain regions, including the prefrontal cortex (PFC), the nucleus accumbens (NAc), the ventral tegmental area (VTA), the amygdala and the hippocampus. Despite the many empirical uncertainties regarding various aspects of the addiction process, some of these changes may involve synapse selection. Other changes may involve non-selective, "Hebbian" mechanisms. Drug addiction not only provides a potential example of neural selection, but it also suggests ways in which neural selection processes can potentially explain some features of addiction. Finally, it illustrates the possibility of conflict between levels of selection. In this case, there may be conflict between the evolutionary function of the dopamine reward system and the ontogenetic function(s) of addiction. Given the empirical uncertainties involved in this example, I cannot claim with certainty that some aspects of addiction involve synapse selection. I merely present this as a plausible candidate for such an explanation.

The central components of the drug addiction process are the mesolimbic and mesocortical dopamine (DA) pathways. DA-carrying neurons in the VTA of the midbrain send excitatory projections to neurons to the NAc (the mesolimbic pathway) and the PFC (the mesocortical pathway). Glutamate carrying neurons in the PFC also send excitatory projections to the NAc. All "drugs of abuse" appear to act on DA-carrying neurons in the VTA in such a way as to promote the release of DA in the NAc and PFC. This release of

dopamine can be associated with a pleasurable sensation, and it also leads to the reinforcement of the behaviors that initiated its release. (For a concise overview of the neuroscience of addiction, see Hyman et al. 2006 and other references below).

"Drugs of abuse" act on the same system that is involved in the reinforcement of rewarding behaviors generally. That is why they are often said to "coopt," "usurp," or "hijack" the normal function of the DA-reward system (Mameli et al. 2011, 414; Hyman et al. 2006, 565; Robbins and Everitt 1999, 567). One difference, however, is that some drugs cause VTA neurons to release a substantially higher amount of dopamine than do "natural" rewards such as food or sex. This has two important consequences. First, the (drug-seeking and drug-taking) behaviors that led to the release of dopamine are reinforced. Second, the addiction process typically produces anhedonia (Goldstein and Volkow 2011, 660), which refers to the loss of pleasure associated with stimuli or behaviors that were formerly rewarding. From a behavioral or motivational perspective, one could say that there is a "competition" between drug-seeking and non-drug-seeking behaviors, such that the former are differentially reinforced over the latter.

From the perspective of the GSE, the differential reinforcement of one behavior (or behavioral disposition) over another suffices to confer a (selected effects) function on the former. This is the case even if it is not known whether the underlying neural mechanisms involve selection processes. In other words, even if it turns out that all of the mechanisms involved at the synapse level in drug addiction are constructive ("Hebbian"), there would still be sufficient warrant for attributing a function to drug-taking *behaviors*.

The function(s) of the drug-taking behavior is to do whatever it did in the past that brought about its differential reinforcement. It is quite possible that this function may differ in different individuals, depending on what those individuals found rewarding about the experience. In all cases, however, one function of the drug-taking behavior is to facilitate the release of DA, since this is how the behavior contributes to its differential reinforcement.

Given that the process of drug addiction brings about novel *behavioral* functions, is it possible that it brings about novel *neural* functions as well? One possibility is that the release of DA in the NAc and PFC facilitates the selective strengthening of synapses between PFC afferents and NAc neurons (Schultz 1998, 15). According to this model, one function of DA release in the NAc is to signal the existence of an unexpected reward and to *selectively reinforce those particular synapses involved in producing the behavior that brought about the reward*. In this model, the release of DA strengthens only those synapses that involve PFC neurons implicated in the rewarding action (ibid).

Is this a synapse selection process or not? In order for this to be a selection process there must be fitness-affecting interactions between the synapses; in order for this to be a competitive process specifically, the strengthening of a synapse associated with the rewarding action must lead to the weakening of synapses associated with other actions. Unfortunately, the precise cellular and molecular mechanisms that facilitate these synaptic changes are not entirely known. If there *are* competitive interactions between synapses, however, this might explain "competitive" interactions on the behavioral level.

That is, it might explain the differential reinforcement of certain (drug-seeking and drug-taking) behaviors by virtue of the differential reinforcement of the synapses that facilitate those behaviors.

Drug addiction may provide an interesting illustration of conflicts between levels of selection. According to some researchers, the evolved function of the DA reward system is to reinforce behaviors that are associated with "natural" rewards such as food and sex (Hyman et al. 2006, 567). From this perspective, drug addiction seems to "coopt," "usurp," or "hijack" those mechanisms. If, however, the process of drug addiction introduces novel functions on the behavioral level, and perhaps the neural level, then this could be described as a conflict between levels of selection. Here, the process of drug addiction, by "coopting" the DA-reward system, has conferred new functions onto that system. From an evolutionary perspective, the function of the DA-reward system may be to facilitate fitness-enhancing behaviors; from a behavioral or neural perspective, it may be to facilitate drug-seeking and drug-taking behaviors which may happen to be fitness reducing.

A selected effects analysis of this sort would also suggest new questions: what other kinds of benefits did the ingestion or effects of "drugs of abuse" serve that led to the differential reinforcement of those behaviors and possibly the neural structures underlying them? There are a lot of reasons that people initially find solace in drugs, from alleviating boredom, to facilitating group membership, to relieving social anxiety, to providing respite from the nauseating mediocrity of day-to-day existence. According to

GSE, these would be amongst the *functions* of the reinforced behaviors and possibly their neural underpinnings. An understanding of the specific benefits that initially led to the reinforcement of these habits in any given individual would be useful for predicting other sorts of cognitive, emotional, or social challenges that the individual faces and it would certainly be useful in designing appropriate interventions.

6 A pluralist conception of brain function

In the foregoing I have attempted to specify precisely the content of the claim that a given synapse, neuron, or group of neurons can undergo "selection," and how this kind of selection is important for extending the etiological theory of functions into the context of neuroscience. As stated in the introduction, one reason that etiological theories have not been applied more systematically to brain function is because of the view that natural selection operating over an evolutionary time scale, and paradigmatically at the level of the individual, is the only kind of function-bestowing selection process that exists. On the surface, this kind of theory is hard to reconcile with the ubiquity of unique and non-heritable synaptic structures that exist within the brain of any given individual – structures, moreover, that perform functions that appear to be evolutionary novel, such as reading, distinguishing models of cars, or playing Tetris. In hindsight, one can see that the inability to appreciate the distinction between neural selection and neural construction is also partly responsible for the fact that etiological theorists have never successfully extended the selected effects theory to the context of neuroscience.

In my view, selected effects theorists ultimately "ceded" the context of neuroscience to consequentialist theorists because of the apparent difficulty in explaining evolutionarily novel functions, particularly when contrasted with the apparent success of consequentialist theories in doing the same. In particular, as noted in the introduction, philosophical reflection on neuroscience has been dominated by the causal role theory associated with the work of Cummins (1975). In Cummins" view, the function of a trait consists, roughly, in the contribution that the trait makes, in tandem with other traits, to producing a phenomenon of interest to a researcher. One reason this theory has become prominent in philosophical reflection on neuroscience is because it dovetails neatly with the fact that many neuroscientists are more preoccupied with structural and functional decompositions of present-day abilities than with reconstructing evolutionary trajectories (though this is by no means universal). Not only has this theory dominated reflection on neuroscience, but it has also been folded into the foundational literature on the "new mechanism" (see, e.g., Bechtel and Richardson 1993, 17; MDC 2000, 6; Craver 2001; forthcoming; Glennan 2002, 127 (fn. 6); Glennan 2005, 456).[11]

Thus, one purpose of the foregoing is to regain some of the ground for the selected effects theory that was unnecessarily ceded. This is not to suggest that the selected effects theory should be seen as replacing or supplanting the causal role view. Clearly, function statements in biology are not always construed as answers to the question, "why is it there?" Moreover, biologists do not always use the notion of function with any historical connotations, but rather, to mark off a special subset of present-day capacities or dispositions of traits (e.g., those that typically contribute to the survival or inclusive

fitness of the organism). Instead of supplanting the causal role view, I endorse a modest pluralism which holds that there are different concepts of function at play in the realm of biology (e.g., Neander 1991, 181; Millikan 2002; Griffiths 2006) – or perhaps, more simply stated, the term "function" is ambiguous.

Some philosophers who have endorsed a pluralist view, however, have suggested a problematic "division of labor" between different explications of function. According to Griffiths (2006), the etiological theory is appropriate to some evolutionary questions because evolutionary theorists often do use the term function with historical connotations. In other contexts, function statements are only associated with questions about present-day capacities of traits, and hence only consequentialist theories are appropriate. Maclaurin and Sterelny (2008, 114) also endorse a similar division when they suggest that, while selected effects functions are appropriate for evolutionary questions, they have no place in ecology, where causal role functions alone are appropriate. My own view is that this "labor-sharing" device unnecessarily limits, or better, marginalizes, the utility of the etiological approach to function.

There is a more productive and conciliatory way of thinking about function pluralism. The etiological theory of function is appropriate in biology not only whenever evolutionary questions are being asked, but whenever *historical* questions are being asked. These questions may either pertain to the evolution of a system or the development of a system over an ontogenetic time scale. In this view, both etiological and consequentialist theories have a natural home in all fields of biology because all

fields of biology ask both historical and non-historical questions. As Brandon points out, in geology, the concept of a mountain ridge is non-historical; the concept of a volcanic mountain is historical (that is, whether or not something is a volcanic mountain depends on its history). Nobody would suggest that the former or the latter concept be expunged from geology (Brandon forthcoming). By the same token, it is not clear why one would dismiss a concept of function that provides a unifying framework for thinking about evolutionarily and ontogenetically derived functions in the brain, and that can illuminate the pervasive historicity of neural structure and activity. After all, the difference between swamp people and real people is their *historicity*; as neuroscience is the science of real people (amongst other organisms), one might expect the concept of brain function to reflect this historicity.

References

Allen C., & Bekoff, M. (1995). Biological function, adaptation, and natural design. *Philosophy of Science*, *62*, 609-622.

Antonini, A., & Stryker, M. P. (1993a). Rapid remodeling of axonal arbors in the visual cortex. *Science*, *260*, 1819-1821.

Antonini, A., & Stryker, M. P. (1993b). Development of individual geniculocortical arbors in cat striate cortex and effects of binocular impulse blockade. *The Journal of Neuroscience*, 13, 3549-3573.

Ayala, F. (1970). Teleological explanations in evolutionary biology. *Philosophy of Science*, *37*, 1-15.

Barlow, H. B. (1988). Neuroscience: a new era? *Nature*, *331*, 571.

Bechtel, W., & Richardson, R. C. (1993). Discovering complexity: decomposition and localization as strategies in scientific research. Princeton, NJ: Princeton University Press.

Black, J. E., & Greenough, W. T. (1986). Induction of pattern in neural structure by experience: implications for cognitive development. In M. E. Lamb, A. L. Brown, & B. Rogoff (Eds.) *Advances in developmental psychology, vol. 4* (pp. 1-50). Hillsdale, N. J.: Lawrence Erlbaum.

Black, J. E., & Greenough, W. T. (1997). How to build a brain: multiple memory systems have evolved and only some of them are constructivist. *Behavioral and Brain Sciences*, *20*(4), 558-559.

Boorse, C. (1976). Wright on functions. *Philosophical review*, *85*, 70-86.

Bothwell, M. (1995). Functional interactions of neurotrophins and neurotrophin receptors. *Annual Review of Neuroscience*, *18*, 223-253.
Bouchard, F. (2008). Causal processes, fitness, and the differential persistence of lineages. *Philosophy of Science*, *75*, 560-570.
Brandon, R.N. (1990) Adaptation and environment. Princeton: Princeton University Press.
Brandon, R. N. (forthcoming). A general case for functional pluralism. In P. Huneman (Ed.) *Functions: selection and mechanisms* (pp. 199-220). Boston: Synthese Library.
Buller, D. J. (2005). Adapting minds: Evolutionary psychology and the persistent quest for human nature. Cambridge, MA: MIT Press.
Buller, D. J., & Hardcastle, V. G. (2000). Evolutionary psychology, meet developmental neurobiology: Against promiscuous modularity. *Brain and Mind*, *1*, 307-325.
Buonomano, D. V., & Merzenich, M. M. (1998). Cortical plasticity: from synapses to maps. *Annual Review of Neuroscience*, *21*, 149-186.
Burnet, F. M. (1959). The clonal selection theory of acquired immunity. Cambridge: Cambridge University Press.
Changeux, J. P. (1985). Neuronal man. New York: Pantheon Books.
Changeux, J. P. (1997). Variation and selection in neural function. *Trends in Neurosciences*, *20*, 291-292.
Changeux, J.-P., & Danchin, A. (1976). Selective stabilization of developing synapses as a mechanism for the specification of neuronal networks. *Nature*, *264*, 705-11.
Changeux, J. –P., Courrège, P., & Danchin, A. (1973). A theory of the epigenesis of neural networks by selective stabilization of synapses. *Proceedings of the National Academy of Sciences of the United States of America, 70*, 2974-2978.
Clarke, P. G. H., & Cowan, W. M. (1975). Ectopic neurons and aberrant development during neural development. *Proceedings of the National Academy of Sciences of the United States of America*, *72,* 4455-4458.
Clarke, P. G. H., & Cowan, W. M. (1976). The development of the isthmo-optic tract in the chick, with special reference to the occurrence and correction of developmental errors in the location and ocnnections of isthmo-optic neurons. *Journal of Comparative Neurology*, *167,* 143-164.
Cohen, S., & Levi-Montalcini, R. (1956). A nerve growth stimulating factor, isolated from snake venom. *Proceedings of the National Academy of Sciences of the United States of America*, *42*, 571-574.
Cowan, W. M. (1973). Neuronal death as a regulative mechanism in the control of cell number in the nervous system. In M. Rockstein (Ed.), *Development and aging in the nervous system*, (pp. 19-41). New York: Academic Press.
Cowan, W. M. (1978). Aspects of neural development. In R. Porter (Ed.), *Neurophysiology III* (pp. 149-191). Baltimore: University Park Press.
Craver, C. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science*, *68*, 53-74.
Craver, C. (forthcoming). Functions and mechanisms in contemporary neuroscience. In P. Huneman (Ed.) *Functions: selection and mechanisms* (pp. 199-220). Boston: Synthese Library.
Crick, F. (1989). Neural edelmanism. *Trends in Neurosciences*, *12*, 240-248.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy*, *72*, 741-765.
Cziko, G. (1995). Without miracles: universal selection theory and the second darwinian revolution. Cambridge: MIT Press.
Darden, L., & Cain, J. A. (1989). Selection type theories. *Philosophy of Science*, *56*, 106-129.
Davies, A. M., Bandtlow, C., Heumann, R., Korsching, S., Hermann, R., & Thoenen, H. (1987). Timing and site of nerve growth factor synthesis in developing skin in relation to innervation and expression of the receptor. *Nature*, *326*, 353-358.
Deppmann, C. D., et. al. 2008. A model for neuronal competition during development. *Science*, *320*, 369-373.
Detwiler, S. R. (1936). Neuroembryology: an experimental study. New York: Macmillan.
Ebendal, T., Olson, L., Seiger, A., & Hedlund, K. –O. (1980). Nerve growth factor in the rat iris. *Nature*, *286*, 25-28.
Edelman, G. (1967). In G. C. Quarton, T. Melnechuk, & F. O. Schmitt (Eds.), *The Neurosciences: A Study* Program (pp. 200-205). New York: Rockefeller University Press.
Edelman, G. (1975). Molecular recognition in the immune and nervous systems. In F. G. Worden, J. P. Swazey, and G. Adelman (Eds.), *The Neurosciences: Paths of Discovery* (pp. 65-74). Cambridge: MIT Press.
Edelman, G. M. (1978). Group selection and phasic reentrant signaling: a theory of higher brain function. In G. M. Edelman, & V. B. Mountcastle (Eds.), *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function* (pp. 51-100). Cambridge, Mass.: MIT Press.
Edelman, G. M. (1987). Neural darwinism: the theory of neuronal group selection. New York: Basic Books.
Edelman, G. M., & Tononi, G. (2001). Consciousness: how matter becomes imagination. London: Penguin.
Elliott, T., & Shadbolt, N. R. (1997). Neurotrophic factors, neural selectionism, and neuronal proliferation. *Behavioral and Brain Sciences*, *20*, 561-562.
Elliott, T., & Shadbolt, N. R. (1998). Competition for neurotrophic factors: Ocular dominance columns. *The Journal of Neuroscience*, *18*, 5850-5858.
Elliott, T., & Shadbolt, N. R. (2002). Multiplicative synaptic normalization and a nonlinear Hebb rule underlie a neurotrophic model of competitive synaptic plasticity. *Neural Computation*, *14*, 1311-1322.
Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience*, *8*, 1481-1489.
Garson, J. (2008). Function and teleology. In S. Sarkar, & A. Plutynski (Eds.), *A Companion to the Philosophy of Biology* (pp. 525-549). Malden, MA: Blackwell.
Garson, J. (2010). Schizophrenia and the dysfunctional brain. *Journal of Cognitive Science*, *11*, 215-246.
Garson, J. (2011). Selected effects functions and causal role functions in the brain: the case for an etiological approach to neuroscience. *Biology & Philosophy*, *26*, 547-565.
Gazzaniga, M. S. (1992). Nature"s mind: the biological roots of thinking, emotions, sexuality, language, and intelligence. New York: Basic Books.

Glennan, S. (2002). Contextual unanimity and the units of selection problem. *Philosophy of Science*, *69*, 118-137.
Glennan, S. (2005). Modeling mechanisms. *Studies in the History and Philosophy of Biological and Biomedical Sciences*, *36*(2), 443-464.
Godfrey-Smith, P. (1992). Indication and adaptation. *Synthese*, *92*, 283-312.
Godfrey-Smith, P. (1994). A modern history theory of functions. *Nous*, *28*, 344-362.
Godfrey-Smith, P. (2007). Conditions for evolution by natural selection. *Journal of Philosophy*, *104*, 489-516.
Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nature Reviews Neuroscience*, *12*, 652-669.
Griffiths, P. E. (1993). Functional analysis and proper function. *British Journal for the Philosophy of Science*, *44*, 409-422.
Griffiths, P. E. (2006). Function, homology, and character individuation. *Philosophy of Science*, *73*, 1-25.
Haier, R. J., Karama, S., Leyba, L., & Jung, R. E. (2009). MRI assessment of cortical thickness and functional activity changes in adolescent girls following three months of practice on a visual-spatial task. *BMC Research Notes*, *2*, 174-180.
Hamburger, V. (1958). Regression versus peripheral control of differentiation in motor hypoplasia. *American Journal of Anatomy*, *102*, 365-410.
Hamburger, V., & Levi-Montalcini, R. (1949). Proliferation, differentiation, and degeneration in the spinal ganglia of the chick embryo under normal and experimental conditions. *Journal of Experimental Zoology*, *111*, 457 – 507.
Harris, A. E., Ermentrout, G. B., & Small, S. L. (1997). A model of ocular dominance column development by competition for trophic factor. *Proceedings of the National Academy of Science USA*, *94*, 9944-9949.
Harris, A. E., Ermentrout, G. B., & Small, S. L. (2000). A model of ocular dominance column development by competition for trophic factor: effects of excess trophic factor with monocular deprivation and effects of antagonist of trophic factor. *Journal of Computational Neuroscience*, *8*, 227-250.
Hebb, D. O. (1949). The organization of behavior. New York: Wiley.
Hodgkin, A. L., & Huxley, A. F. (1939). Action potentials recorded from inside a nerve fiber. *Nature*, *144*, 710-711.
Hollyday, M., & Hamburger, V. (1976). Reduction in naturally occurring motor neuron loss by enlargement of the periphery. *Journal of Comparative Neurology*, *170*, 311-320.
Huang, E. J., & Reichardt, L. F. (2001). Neurotrophins: roles in neuronal development and function. *Annual Review of Neuroscience*, *24*, 677-736.
Hubel, D. H., & Wiesel, T. N. (1965). Binocular interaction in striate cortex of kittens reared with artificial squint. *Journal of Neurophysiology*, *28*, 1041-1059.
Hubel, D. H., & Wiesel, T. N. (1972). Laminar and columnar distribution of geniculo-cortical fibers in the macaque monkey. *Journal of Comparative Neurology*, *146*, 421-450.
Hubel, D. H., Wiesel, T. N., & LeVay, S. (1977). Plasticity of ocular dominance columns in monkey striate cortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *278,* 377-409.

Hull, D. L., Langman, R. E., & Glenn, S. S. (2001). A general account of selection: biology, immunology and behavior. *Behavioral and Brain Sciences*, *24*, 511–527.
Hyman, S. E., Malenka, R. C., & Nestler, E. J. (2006). Neural mechanisms of addiction: the role of reward-related learning and memory. *Annual Review of Neuroscience*, *29*, 565-598.
Jerne, N. K. (1955). The natural-selection theory of antibody formation. *Proceedings of the National Academy of Sciences of the United States of America*, *41*(11), 849-857.
Jerne, N. K. (1967). Antibodies and learning: selection vs. instruction. In G. C. Quarton, T. Melnechuk, & F. O. Schmitt (Eds.), *The Neurosciences: A Study* Program (pp. 200-205). New York: Rockefeller University Press.
Johnson Jr., E. M., & Deckworth, T. L. (1993). Molecular mechanisms of developmental neuronal death. *Annual Review of Neuroscience*, *16*, 31-46.
Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). Principles of neural science. 4th ed. New York: McGraw Hill
Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*, *183*, 339-373.
Katz, L.C., & Shatz, J. C. (1996). Synaptic activity and the construction of cortical circuits. *Science*, *234*, 1133-1138.
Kingsbury, J. (2008). Learning and selection. *Biology and Philosophy*, *23*, 493-507.
Lederberg, J. (1959). Genes and antibodies. *Science*, *129*, 1649-1653.
LeDoux, J. (2002). Synaptic self: how our brains become who we are. New York: Penguin.
Lennox, J. G., & Wilson, B. E. (1994). Natural selection and the struggle for existence. *Studies in the History and Philosophy of Science*, *25*, 65-80.
Levi-Montalcini, R., & Cohen, S. (1960). Effects of the extract of the mouse submaxillary glands on the sympathetic system of mammals. *Annals of the New York Academy of Sciences*, *85*, 324-341.
Lewens, T. (2007). Functions. In M. Matthen, C. Stevens, D. M. Gabbay, P. Thagard, & J. Woods (Eds.), *Philosophy of biology* (pp. 525-547). Amsterdam: Elsevier.
Lewontin. (1970). The units of selection. *Annual Review of Ecology and Systematics*, *1*, 1-18.
Lichtman, J. W, Burden, S. J., Culican, S. M., & Wong, R. O. L. (1999). Synapse formation and elimination. In Zigmond, M. J, Bloom, F. E, Landis, S. C., Roberts, J. L., and Squire, L. R. (Eds.), *Fundamental Neuroscience* (pp. 547-580). San Diego: Academic Press.
Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*, 1-25.
Maclaurin, J., & Sterelny, K. (2008). *What is biodiversity?* Chicago: University of Chicago.
Mameli, M., Bellone, C., Brown, M. T. C., & Lüscher, C. (2011). Cocaine inverts rules for synaptic plasticity of glutamate transmission in the ventral tegmental area. *Nature Neuroscience*, *14*, 414-416.
McDowell, J. J. (2009). Behavioral and neural darwinism: selectionist function and mechanism in adaptive behavior dynamics. *Behavioural Processes*, *84*(1), 358-365.

Meyer, R. L. (1998). Roger Sperry and his chemoaffinity hypothesis. *Neuropsychologia*, *36*, 957-980.
Meyer, R. L., & Sperry, R. (1976). Retinotectal specificity: chemoaffinity theory. In Gottlieb, G. (Ed.), *Studies on the Development of Behavior and the Nervous System, Vol. 3: Neural and Behavioral Specificity* (pp. 111-149). New York: Academic Press.
Millikan, R. G. (1984). Language, thought, and other biological categories. Cambridge: MIT Press.
Millikan, R. G. (1989). In defense of proper functions. *Philosophy of Science*, *56*, 288-302.
Millikan, R. G. (2002). Biofunctions: two paradigms. In Ariew, A., Cummins, R., & Perlman, M. (Eds.), *Functions: New Essays in the Philosophy of Psychology and Biology* (pp. 113-143). Oxford: Oxford University Press.
Millstein, R. L. (2009). Populations as individuals. *Biological Theory*, *4*, 267-273.
Mitchell, S. D. (1995). Function, fitness, and disposition. *Biology and Philosophy*, *10*, 39-54.
Neander, K. (1983). Abnormal psychobiology. Dissertation, La Trobe.
Neander, K. (1991). Functions as selected effects: the conceptual analyst"s defense. *Philosophy of Science*, *58*, 168-184.
Neander, K. (1999). Fitness and the fate of unicorns. In V. G. Hardcastle (Ed.), *Where Biology Meets Psychology* (pp. 3-26). Cambridge, MA: MIT Press.
Neander, K. (2008). Teleological theories of mental content: Can Darwin solve the problem of intentionality? In M. Ruse (Ed.), *The Oxford handbook of philosophy of biology* (pp. 381-409). Oxford: Oxford University Press.
Okasha, S. (2003). Does the concept of "clade selection" make sense? *Philosophy of Science*, *70*, 739-751.
Oppenheim, R. W. (1989). The neurotrophic theory and naturally occurring motoneuron death. *Trends in Neurosciences*, *12*, 252-255.
Oppenheim, R. W. (1991). Cell death during development of the nervous system. *Annual Review of Neuroscience*, *14*, 453-501.
Papineau, D. (1987). Reality and representation. Oxford: Blackwell.
Papineau, D. (1993). Philosophical naturalism. Oxford: Blackwell.
Papineau, D. (1995). Mental disorder, illness and biological disfunction. In Griffiths, A. P. (Ed.), *Philosophy, psychology and psychiatry* (pp. 73-82). Cambridge: Cambridge University Press.
Pettmann, & Henderson. (1998). Neuronal cell death. *Neuron*, *20*, 653-647.
Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese*, *183*, 283-311.
Price, D. J., Jarman, A. P., Mason, J. O., & Kind, P. C. 2011. Building brains: An introduction to neural development. Chichester: Wiley-Blackwell.
Porges, S. W., & Carter, C. S. (2011). Mechanisms, mediators, and adaptive consequences of caregiving. In S. L. Brown, R. M. Brown, & L. A. Penner (Eds.), *Moving beyond self-interest: Perspectives from evolutionary biology, neuroscience, and the social sciences* (pp. 53-71). Oxford: Oxford University Press.

Purves, D. (1988). A new theory of brain function. *The Quarterly Review of Biology*, *63*, 202-204.
Purves, D. (1994). Neural activity and the growth of the brain. Cambridge: Cambridge University Press.
Purves, D., White, L. E., & Riddle, D. R. (1996). Is neural development darwinian? *Trends in Neuroscience*, *19*, 460-464.
Quartz, S. R., & Sejnowski, T. J. (1997). The neural basis of cognitive development: a constructivist manifesto. *Behavioral and Brain Sciences*, *20*, 537-596.
Rakic, P. (1976). Prenatal genesis of connections subserving ocular dominance in the rhesus monkey. *Nature*, *261*, 467-471.
Robbins, T. W., & Everitt, B. J. (1999). Drug addiction: bad habits add up. *Nature*, *398*, 567-570.
Schaffner, K. (1993). Discovery and explanation in biology and medicine. Chicago: University of Chicago Press.
Schlaggar, B. L., & McCandliss, B. D. (2007). Development of neural systems for reading. *Annual Review of Neuroscience*, *30*, 475-503.
Scholes, J. (1979). Nerve fiber topography in the retinal projection to the tectum. *Nature*, *278*, 620-624.
Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*, 1-27.
Schwartz, P. H. (1999). Proper function and recent selection. *Philosophy of Science*, *66*, S210-S222.
Skinner, B. F. (1953). Science and human behavior. New York: The Free Press.
Skinner, B. F. (1981). Selection by consequences. *Science*, *213*, 501-504.
Sober, E. (1984). The nature of selection. MIT Press: Cambridge.
Sober, E., & and Wilson, D. S. (1999). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Harvard University Press: Cambridge, MA.
Sperry, R. (1951). Mechanisms of neural maturation. In Stevens, S. S. (Ed.), *Handbook of Experimental Psychology* (pp. 236-280). New York: Wiley.
Sperry, R. (1963). Chemoaffinity in the orderly growth of nerve fiber patterns of connections. *Proceedings of the National Academy of Sciences of the United States of America*, *50*, 703-710.
Sporns, O. (1997a). Variation and selection in neural function. *Trends in Neurosciences*, *20*, 291.
Sporns, O. (1997b). Deconstructing neural constructivism. *Behavioral and Brain Sciences*, *20*, 576-577.
Van Ooyen, A. (2011). Using theoretical models to analyze neural development. *Nature Reviews Neuroscience*, *12*, 311-326.
Walicke, P. A. (1989). Novel neurotrophic factors, receptors, and oncogenes. *Annual Review of Neuroscience*, *12*, 103-126.
Walsh, D. M., & Ariew, A. (1996). A taxonomy of functions. *Canadian Journal of Philosophy*, *26*, 493-514.
Weiskopf, D. A. (2011). Models and mechanisms in psychological explanation. *Synthese*, *183*, 313-338.
Wiesel, T. N., & Hubel, D. H. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of Neurophysiology*, *26*, 1003-1017.

Wilson, D. S. (1975). A theory of group selection. *Proceedings of the National Academy of Sciences USA*, *72*, 143-146.
Wimsatt, W. C. (1972). Teleology and the logical structure of function statements. *Studies in the History and Philosophy of Science*, *3*, 1-80.
Wouters, A. (2003). Four notions of biological function. *Studies in the History and Philosophy of Biological and Biomedical Sciences*, *34*, 633-668.
Wright, L. (1973). Functions. *Philosophical Review*, *82*, 139-168.
Wright, L. (1976). Teleological explanations: an etiological analysis of goals and functions. Berkeley: University of California Press.
Young, J. Z. (1936). The structure of nerve fibres in cephalopods and Crustacea. *Proceedings of the Royal Society of London B*, *121*, 319-337.
Young, J. Z. (1964). A model of the brain. Oxford: Clarendon Press.
Yuan, J., & Horvitz, H. R. (1990). The *Caenorhabditis elegans* genes *ced-3* and *ced-4* act cell autonomously to cause programmed cell death. *Developmental Biology*, *138*, 33-41.

---

[1] Specifically, why there is a non-zero frequency of stingers amongst honeybees; see Griffiths 1993, 415 for an important discussion of what, precisely, function statements purport to explain.

[2] Wimsatt (1972, 14), similarly, noted that trial-and-error learning is a kind of selection process.

[3] Not all philosophers would accept that natural selection *as such* operates over entities that do not reproduce. This is because, in their view, a *sine qua non* of natural selection in the evolutionary context is the existence of a population of reproducing entities (e.g., Okasha 2003; cf. Godfrey-Smith 2007 for an overview of definitions of "natural selection" and a description of the role of "reproduction," and Bouchard 2008 for a dissenting view). This debate is of marginal relevance to my view of function. If one accepts that natural selection per se requires a population of reproducing entities, the only consequence for my view is that neural selection is not a form of natural selection. But nothing of importance rides on this. All that matters is that neural selection and natural selection (however one chooses to define the latter) can be viewed as two different types of "selection processes" construed at a fairly abstract level of description, that is, in such a way as to satisfy GSE below.

[4] The point is argued in Lennox and Wilson (1994, 70); the latter is B. E. Wilson and not D. S. Wilson.

[5] Of course, one may choose to amend this simple formulation with a clause that restricts functions to effects that were selected for in the recent past, in the manner of Griffiths (1993), Godfrey-Smith (1994), or Schwartz (1999). The latter holds that the functional activity must have been selected for at some stage in its evolution, and that it must have recently contributed to survival and reproduction (without necessarily having been selected for recently).

[6] This also excludes the differential reproduction of a sequence of non-functional DNA by genetic drift from having a function: although it undergoes differential reproduction, it does not in any obvious sense "contribute" to its differential reproduction (Godfrey-Smith 2007; Section 6).

[7] I am grateful to an anonymous reviewer for pointing this out.
[8] Interestingly, Young also discovered the utility of the giant axon of the squid for neurophysiology (Young 1936), which was famously used by Hodgkin and Huxley (1939) in their Nobel-prize winning research.
[9] I thank an anonymous referee for this observation.
[10] Even Purves et al. (1996, 462), in a very critical paper, regard ocular dominance plasticity as a paradigm of neural selection. They just question the prevalence of this mechanism in the formation of other neural structures.
[11] Though Bechtel and Richardson do not use the term "causal role" in their foundational contribution to the literature, they characterize the "function" of a part of a system in terms of its contribution, in tandem with the other parts, to the "behavior" of the system as a whole (Bechtel and Richardson 1993, 17).