Biol Philos (2011) 26:547–565
DOI 10.1007/s10539-011-9262-6

Selected effects and causal role functions in the brain: the case for an etiological approach to neuroscience

Justin Garson

Received: 17 October 2010 / Accepted: 7 March 2011 / Published online: 18 March 2011

Abstract Despite the voluminous literature on biological functions produced over the last 40 years, few philosophers have studied the concept of function as it is used in neuroscience. Recently, Craver (forthcoming; also see Craver 2001) defended the causal role theory against the selected effects theory as the most appropriate theory of function for neuroscience. The following argues that though neuroscientists do study causal role functions, the scope of that theory is not as universal as claimed. Despite the strong *prima facie* superiority of the causal role theory, the selected effects theory (when properly developed) can handle many cases from neuroscience with equal facility. It argues this by presenting a new theory of function that generalizes the notion of a 'selection process' to include processes such as neural selection, antibody selection, and some forms of learning—that is, to include structures that have been differentially retained as well as those that have been differentially reproduced. This view, called the generalized selected effects theory of function, will be defended from criticism and distinguished from similar views in the literature.

Keywords Teleology · Function · Neuroscience · Mechanism · Selection · Neural Darwinism

Introduction

The last 40 years have seen an explosion of philosophical work on the subject of naturalized teleology, and specifically on the concept of biological function (see Garson 2008 for an overview). Despite substantial theoretical differences between

J. Garson (✉)
Department of Philosophy, Hunter College of the City University of New York,
695 Park Avenue, New York, NY 10065, USA
e-mail: justin.garson@hunter.cuny.edu

the accounts, most philosophers have relied on the conceptual apparatus of evolutionary theory to understand function. Specifically, function is typically tied to the fitness or reproductive capacity of organisms. Take a typical function statement such as, "the function of the heart is to circulate blood." Depending on one's persuasion, one will either analyze this as a historical claim (e.g., "the heart was *selected for by natural selection* for circulating blood") or as a claim about a subset of current capacities or dispositions of the heart (e.g., "blood circulation is the heart's current species-typical contribution to survival and reproduction"). Theories within the former group are known as *etiological*, those in the latter, *consequentialist*—though some have opted for a mixed theory that incorporates elements of both (e.g., Wright 1973, 161; Kitcher 1993, 387; Walsh and Ariew 1996, 501; Matteo et al. 2009, 828).

One consequentialist theory that departs from evolutionary considerations is the causal role (CR) account, which has been defended for several decades (see Cummins 1975; Prior 1985; Amundson and Lauder 1994; Hardcastle 1999; Davies 2001; Craver 2001, forthcoming). CR examines the role of a trait within a complex, hierarchically organized system. In particular, it examines the way that trait interacts with others to yield a complex capacity of that system. Because, in principle, a trait can have an infinite number of functions (depending on which system and systemic capacity is under consideration) CR theorists typically relativize the function of a trait to the interests of the investigator. One consequence of CR is that evolutionary considerations are not required to ascribe a function to a trait (unless the reproductive capacity of the organism happens to be the systemic capacity under consideration). It is also neutral between biological and artifact functions. It applies just as easily to mousetraps as to memory.

Probably the majority of philosophers of science today accept the etiological approach to function, and more specifically, a version known as the 'selected effects' (SE) theory (e.g., Neander 1983; 1991; Millikan 1984; 1989; Sober 1984, 208; Brandon 1990, 184–189; Griffiths 1993; Godfrey-Smith 1994; Mitchell 1995; Allen and Bekoff 1995; Schwartz 1999). At its most unqualified level, SE holds that *being selected for* is necessary and sufficient for possession of a function. One reason, however, for the ready acceptance of SE is that it was largely developed using 'comfortable' and familiar function ascriptions, such as "the function of the heart is to circulate blood;" "the function of the kidney is to extract wastes from the blood;" and so on. These cases support SE because they seem to possess an evolved 'purpose' shaped by natural selection and the individual organism is the beneficiary of this purpose.

However, SE is not as unequivocally successful when applied to more unusual biological cases. For example, Godfrey-Smith (1994, 348) uses the example of segregation distorter genes to test intuitions about function. Segregation distorter genes disrupt normal meiosis and thereby guarantee their own disproportionate representation in the sex cells. On the one hand, segregation distorter genes undergo a type of selection insofar as their activity guarantees an increase in their frequency over the generations, at least in the short run. However, since the organism *per se* is not the beneficiary of the selection process (and in fact is disadvantaged by the segregation distorter genes) it seems counterintuitive to assign them the 'function'

of disrupting meiosis. Hence, challenging test cases for SE are those in which: i) the functional trait does not seem to have an evolved function at all (e.g., it is evolutionarily unprecedented or novel) or; ii) it is selected for but does not 'benefit' the organism.

Neuroscience, like molecular biology, offers an exceptional testing ground for competing theories of function. Of course, the brain provides plenty of 'comfortable' and familiar examples that conform to the etiological paradigm, such as the function of the amygdala to produce fear or the function of the visual cortex to produce visual representations. However, it also presents strong difficulties for the etiological theory. Specifically, neuroscientists routinely assign functions to neural structures that: i) are evolutionarily unprecedented and; ii) do not 'benefit' the individual organism in a biologically meaningful way—such as reading ability.

The prevalence of such cases provides a strong argument for the *prima facie* superiority of CR in the context of neuroscience. There are two advantages. First, CR appears to be more consistent with neuroscientific practice, which is more preoccupied with structural and functional decompositions of complex abilities than with speculation about evolutionary histories. Secondly, CR can assign functions to traits regardless of whether those traits evolved by natural selection or whether they 'benefit' the individual organism (Craver forthcoming).

However, the strong *prima facie* superiority of CR is put into question once SE is properly developed. The problem is that many proponents and detractors of SE have suggested that, according to that theory, natural selection operating at the level of the individual organism, and over an evolutionary time scale, is the only process relevant for the ascription of biological functions (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). As a consequence, it is widely assumed that SE cannot ascribe functions in any direct manner to evolutionarily novel traits or to traits that do not benefit the individual.

However, the concept of a 'selection process' is much more general than natural selection operating at the level of the individual over an evolutionary time frame. In fact, selection processes *as such* are ubiquitous in the biological world. Not only does selection act at multiple levels of the biological hierarchy but in very different domains. For example, there are *neural* selection processes, antibody selection processes, and selection processes underlying some types of learning, such as operant conditioning (e.g., Darden and Cain 1989; Cziko 1995; Hull et al. 2001) All of these selection processes operate over the lifetime of the individual and not necessarily intergenerationally. Insofar as they constitute genuine selection processes they can give rise to novel biological functions. SE should in no way be restricted to natural selection acting over an evolutionary time scale, but should be generalized to include these other selection processes.

This broadened view will be called the *generalized selected effects theory* of function (GSE) and explicated and defended below (also see Garson 2010). GSE is best understood as a version of SE that rests on a specific interpretation of what it is to be 'selected for.' Although GSE is a novel theory of function, it is based on similar theories that have been presented (e.g., Wimsatt 1972; Millikan 1984; Papineau 1993). In Sect. Response to criticism and relation to other views, this view

will be carefully distinguished from similar views. Once SE is properly generalized it can handle the problematic cases from neuroscience as well as the causal role theory does. This suggests that the scope of SE is much broader than Craver allows.

The following is divided into five sections. Section Fear responses, spelling rules, and Tetris modules introduces three 'test cases' from neuroscience. It presents an argument for the *prima facie* superiority of CR. Section The generalized selected effects theory of function presents the generalized selected effects theory (GSE) and shows how it can resolve the test cases. Section Response to criticism and relation to other views defends GSE from criticism and distinguishes it from related philosophical attempts to define 'function' in terms of a generalized notion of selection. The final section suggests directions for future research.

Fear responses, spelling rules, and Tetris modules

In this section, three 'test cases' from neuroscience will be presented. The first, the function of the amygdala in producing fear responses, makes a strong case for the usefulness of SE in neuroscience. The second case, the function of a brain region involved in recognizing English spelling rules, is more difficult for SE because it is evolutionarily unprecedented (that is, learned), and has dubious benefits for the individual. The third (a brain region that appears to be specialized for playing Tetris) is similarly challenging to SE because it is both evolutionarily unprecedented and has no clear benefit to the individual. In all three cases CR performs well, which supports its *prima facie* superiority.

The amygdala and the normal fear response

The amygdala is a pair of subcortical nuclei located within the temporal lobe. It has been found to play an important role in the regulation of emotion, and its chief function seems to be fear conditioning. In particular, it is involved in recognizing fearful stimuli, creating new associations between fear and novel stimuli, and triggering appropriate physiological responses. There is also evidence that structural and functional abnormalities of the amygdala (e.g., reduced amygdala volume or lowered amygdala response) can underwrite psychopathy, which appears to involve an inability to empathize with other people (Blair 2003).

The claim that the function of the amygdala is to regulate fear sits naturally with SE. First, although it is always dangerous to speculate about the lifestyles of our Pleistocene ancestors (e.g., Gould and Lewontin 1979), it seems inherently plausible that the ability to experience a moderate level of fear would bestow a fitness advantage upon our evolutionary ancestors over those with no such capacity or who experience crippling fear. For example, some monkeys with temporal lobe lesions that include the amygdala have lost their fear of snakes, which can be disadvantageous. If a heritable neural structure is selected for regulating fear, it comes to have the *function* of regulating fear. Consequently, the function ascription doubles as an explanatory account of that structure. Secondly, it also makes sense of how it is that the amygdala can 'malfunction' or become 'dysfunctional.' Setting aside some

niceties, if reduced amygdala volume or structural abnormalities create a loss of fear or empathy, the amygdala can be said to be unable to perform its selected function, or 'dysfunctional.' SE preserves widespread intuitions about the explanatory and normative dimensions of function ascriptions. In fact, these are amongst its core virtues (e.g., Millikan 1989, 294; Neander 1991, 181; Lewens 2007, 533).

How does CR fare with respect to the same example? In one sense it performs faultlessly. CR would draw attention to two different aspects of amygdala function—on the one hand, its decomposition into component parts and processes, and on the other hand, its integration with other parts and processes in the regulation of the emotional life of human beings (Craver 2001 refers to these as 'constitutive' and 'contextual' explanations respectively). With respect to the capacity for *fear regulation* (as opposed to, say, the capacity for empathy), the amygdala is primarily responsible for recognizing fearful stimuli and triggering appropriate responses, though this function can be modulated by inputs from the hippocampus, sensory cortex, and prefrontal cortex. If one knows the specific contribution of the amygdala to the regulation of fear, as well as its mechanistic interactions with other brain systems that allow it to carry out this role, one can determine its *function*.

However, if a different capacity is chosen (for example, the contribution of the amygdala to empathy) a different distribution of functions necessarily results. For example, from the perspective of its contribution to empathy, the amygdala seems to play a crucial role in interpreting others' facial expressions. One of the well-known features of CR is the manner in which it relativizes the function of an entity to the interests of the investigator, a feature that Craver forthcoming refers to as 'perspectivalism.' There are no functions *proper*, that is, independently of a specific explanatory framework and a specific systemic capacity that one is interested in.

Brain region for orthographic regularities

Recent neuroimaging studies have identified a region in the extrastriate (visual) cortex that is differentially activated in the presence of words that conform to the spelling rules of English (or that exhibit 'orthographic regularity'; Petersen et al. 1990). For example, it will respond not only to common English words such as 'wheel,' but also 'tweal,' which exhibit orthographic regularity. This area, now known as the visual word form area (VWFA), is located in the left occipitotemporal region in a structure called the fusiform gyrus. Recent attempts to understand the mechanisms underlying reading development have given the VWFA an important role (see Schlaggar and McChandliss 2007 for a review).

Clearly this ability is evolutionarily unprecedented—the English language has not existed for long enough for selection to promote a novel brain region specialized for the analysis of visual words. Consequently, SE would, on first glance, deny a function to this important structure. Moreover, it is debatable that reading ability actually enhances the fitness of the organism. Some studies suggest that educational attainment generally (of which reading ability is a part) is associated with lowered fertility (Mathews and Ventura 1997). Consequently, this example not only calls into question the connection between function and selection but the connection between function and present-day contributions to fitness as embodied in

'propensity' theories of function (e.g., Ruse 1971, 91–92; Bigelow and Pargetter 1987, 192; Wouters 2003).

Even if the effect of reading ability on fitness is debatable, other examples from neuroscience suffice to show that neuroscientists do not restrict function statements to traits that somehow 'benefit' the organism, whether conceived in terms of a present-day contribution to fitness or even a general organismic 'good' (e.g., Bedau 1992, 794; McLaughlin 2001, 168). For example, in 2009, researchers at University of New Mexico studied the effect of playing Tetris on the brain (Haier et al. 2009). What they found is that individuals who play Tetris on a regular basis show significant increases in cortical thickness in certain brain regions. One of the brain areas showing the most significant changes is BA22 (ventral temporal lobe), which is associated with the integration of multisensory information (e.g., visual, spatial, tactile, and auditory). The researchers interpreted these changes at BA22 as having the function of promoting the specific multimodal integration capacities *necessary to succeed at Tetris* (ibid., 179). They specifically state that there is no current evidence these capacities enhance other abilities.

Clearly, this study poses a serious challenge to SE—after all, the brain changes in question have not been selected for by natural selection to support Tetris play, yet the brain region seems to have taken on a new function. Even worse, if one makes the reasonable assumption that playing Tetris does not contribute to the current-day survival or reproductive capacity of the individual, this example divorces function entirely from evolutionary considerations. Not only SE but any theory that defines function in terms of current-day benefit or adaptiveness would have to be rejected.

Given the apparent failure of the selected effects theory, how does CR handle the case of the VWFA? The example does not present any principled difficulties to the view. The CR theorist would first select the relevant capacity, such as the capacity to *read and understand the visual word*. He or she would then examine the contribution that this structure makes to that ability. A similar type of analysis would be given for the brain changes associated with the ability to play Tetris. There are many empirical and conceptual details of such an account that await resolution. Advocates of CR have argued that their account actually *facilitates* the discovery of the relevant mechanisms, for example, by synthesizing the existing body of knowledge and thereby drawing attention to gaps and lacunae within it (Robins and Craver 2009, 56) or by allowing one to temporarily and strategically 'bracket' certain problematic phenomena at one hierarchical level to focus exclusively on another (Davies 2001, 87–88). The facility with which CR handles the three cases constitutes a strong *prima facie* case for its superiority in the context of neuroscience. This apparent superiority, however, will be brought into question.

The generalized selected effects theory of function

A central point about SE must be addressed before returning to the test cases from neuroscience. The core problem is that the concept of selection deployed in SE is often interpreted to refer to natural selection operating over an evolutionary time scale (see references in Sect. Introduction). However, selection processes as such are

ubiquitous in the natural world. Once one understands the generality of the concept of a selection process, one can resolve the test cases. In addition to natural selection, there are at least three other well-defined selection processes in the biological world: neural selection, antibody selection, and selection processes that underlie certain forms of learning such as operant conditioning or trial-and-error learning. This observation has often been made but rarely applied to the function debate.

First there are 'neural selection' processes that operate throughout the lifetime of the individual (Changeux and Danchin 1976; Edelman 1987). Often, there are several different neural structures within an individual's brain or nervous system that perform a similar task, with varying levels of effectiveness. A 'neural selection' process is one that causes certain of those structures to be retained and others eliminated over a period of days, weeks, months, or years, because of their differing levels of effectiveness. Simply stated, individual *neurons*, as well as individual *organisms*, are capable of undergoing a type of 'competition' for the resources needed to sustain their form. This constitutes a form of 'neural selectionism' or 'neural Darwinism'—a selection process that operates at the level of individual synaptic structures, neurons, and perhaps even whole groups of neurons, over the lifetime of the individual. Sometimes this process is referred to as 'synaptic pruning' to describe the way that neural connections are gradually eliminated. Neural selection has been shown to play a role in the formation of diverse neural structures such as the neuromuscular junction, ocular dominance columns in the visual cortex, brain regions underlying filial imprinting, and the structure of the olfactory system in mammals (see Wong and Lichtman 2002 for a review).

A specific example of neural selection will be developed by way of illustration. Initial evidence for neural selection came from studies of the neuromuscular junction in mammals (See Fig. 1). At birth, each muscle fiber is typically connected to *several* different motor neurons. Over the course of several weeks, some of those neurons retract, and a one-to-one pattern of connections emerges (see Purves and Lichtman 1980 for an early review of this research). It has been proposed that a competitive mechanism is involved in this eliminative process. One such mechanism would involve the production of some trophic or nutritive material that is synthesized or made available in limited quantities by the muscle, and which is taken up by the innervating neurons and is necessary for the maintenance of the synapse (Brown et al. 1976). The crucial idea is that in some parts of the brain and nervous system, the formation of mature synaptic structures takes place by eliminating or pruning an initially hyperabundant network of neurons, rather than the mere growth of new synapses on an as-needed basis.[1]

The analogy between neural selection and natural selection is clear. In the case of neural selection, one can loosely speak of a *population* of synapses all of which rely on a specific resource for their persistence (for example, a population of neurons that innervate the same target neuron). This population exhibits *variation*, both in spatial position and, most likely, in frequency of firing. Finally, some of these synapses, by virtue of this variation, are *differentially retained* over others. As noted

[1] The difference between selectionist and non-selectionist accounts of synaptic structure formation will be described in the following section.

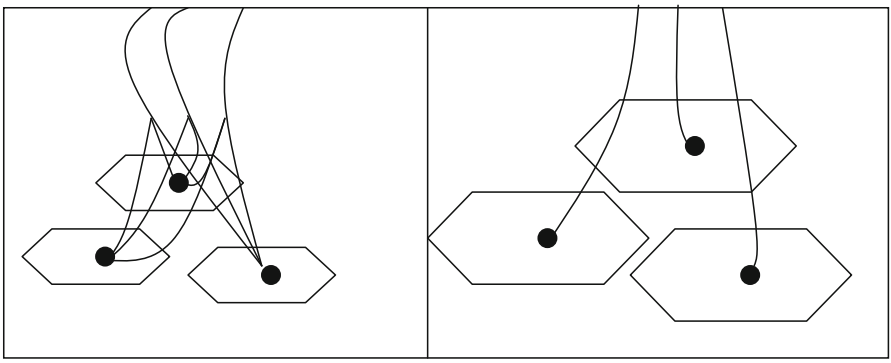

Fig. 1 Innervation of skeletal muscle of newborn rats. The *first panel* depicts the multiple innervation of muscle fibers by motor neurons; the *second panel* depicts the one-to-one pattern of connections that emerges by two weeks after birth. Redrawn from Purves and Lichtman (1980, 155)

above, this may be due to their ability to exploit a trophic substance made available by the target neuron. This would represent a more proximal consequence. Another possibility is that one synapse produces an organismic-level behavior that leads to a reward, and is thereby differentially reinforced by dopamine-dependent reward signals (Schultz 1998). This would represent a more distal consequence.[2]

Just as there are levels of selection in the evolutionary context, there are levels of selection in the neural context as well. Most discussions in the neurobiological literature consider selection at the level of the individual synapse. A higher level of selection would be selection at the level of entire *neurons*. A well-known developmental phenomenon called *neural cell death*, or apoptosis, appears to involve a selection process in which entire neurons (rather than synapses) 'compete' for a limited field of innervation or for a limited number of trophic resources (e.g., Pettmann and Henderson 1998). At a higher level, one would have selection at the level of groups, or 'neural group selection.' According to this view, one outcome of normal developmental processes is the construction of large repertoires of neural groups. Each group in the repertoire exhibits a different internal pattern of connectivity but responds in various degrees to the same stimulus pattern. 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 those of others; e.g., Edelman 1978, 64–65). As yet, there is little direct neurobiological evidence for the existence of neural group selection. In other words, while the evidence for selection at the level of synapses and individual neurons is not disputable, evidence for selection at the level of entire groups of neurons has been questioned (e.g., Crick 1989).

Another type of selection process is antibody selection. Immunologists are familiar with the clonal theory of antibody production, according to which certain

[2] Not all biologists, nor philosophers of biology, would accept neural selection as a type of 'natural selection.' This is because, in their view, reproduction is a *sine qua non* of natural selection (e.g., Okasha 2003; see Bouchard 2008 for an opposing view and Godfrey-Smith 2007 for an overview of definitions of 'natural selection' and a description of the role of 'reproduction').

antibodies are differentially reproduced within the organism as a consequence of exposure to the corresponding antigen (Burnet 1959; Rajewsky 1996). In principle, this can give rise to novel (i.e., evolutionarily unprecedented) functions. Suppose, for example, a random mutation occurs at one of the genes that cause antibody diversity, giving rise to a novel antibody. According to the standard selected effects theory, this antibody would not have a biological function since it was not specifically selected for by natural selection. However, suppose that the antibody becomes exposed to its corresponding antigen, as a consequence of which it differentially replicates within the bloodstream. After having undergone this selection process it would acquire a novel function.

Finally, some forms of learning, most obviously that mediated by operant conditioning or trial-and-error learning, can be modeled as selection processes in which certain behavioral dispositions are differentially retained, and others differentially extinguished, by virtue of their relative performance on a common task. McDowell 2009 gives a recent defense of what he calls 'behavioral Darwinism' and provides a formal comparison with neural Darwinism.[3] Natural selection, neural selection, antibody selection, and some forms of learning constitute selection processes that promote either the differential reproduction or the differential retention of certain structures. It is important to note that not all forms of learning can be construed as 'selection processes,' such as learning by imitation (i.e., modeling; Kingsbury 2008). This point will be discussed in further detail in the next section.

These observations can be used to formulate a new version of the selected effects theory of function, the *generalized selected effects* theory (GSE):

> GSE: The function of a trait consists in that activity that historically contributed to its being differentially reproduced *or* differentially retained within a biological system.

GSE could be more eloquently glossed by saying that the function of a trait consists in that activity which it was 'selected for.' However, such a view could easily be confused with that in which the function of a trait is equivalent to the activity that it was selected for by *natural selection,* which would be overly restrictive. 'Differential reproduction' is intended to capture traits that have been selected for by natural selection operating over an evolutionary time scale, as well as antibodies that are differentially replicated throughout the bloodstream. However, 'differential retention' is intended to capture traits that have been selected for by neural selection or behavioral selection, because neural structures and behavioral dispositions are not necessarily 'reproduced' but can be retained over others. As indicated earlier, GSE is best construed as a version of SE that rests on a specific interepretation of what it is to be 'selected for.' Different interpretations would give rise to different versions of SE. Thus, if one has shown that GSE is compatible with neuroscience, one has shown that SE is compatible with neuroscience.

[3] Mace 1949 (1935) was one of the first to point out a conceptual link between operant conditioning and teleology, followed by Scheffler 1958.

How does GSE handle the two refractory test cases from neuroscience? In brief, in order to show that the VWFA or the 'Tetris module' has a novel function, one could show that a neural selection process had formed the relevant structures—or at least significant parts of those structures. The sorts of speculative evolutionary histories that sometimes underwrite ascriptions of etiological functions will not be useful here. What will be useful is either detailed etiological investigations into the origin of those structures, or general neurodevelopmental models that suggest that most types of complex learning involve a form of selection. The little that is known about the formation of the relevant structures, however, already provides some basis for inferring neural selection.

It has long been established that neural selection is responsible for *some* structures of the visual cortex, such as the formation of ocular dominance columns (e.g., Wiesel and Hubel 1963; Antonini and Stryker 1993). Certain linguistic capacities have also been explained in selectionist terms. Deacon 1989 explains the 'cooption' of vocalization in humans by the prefrontal cortex (rather than the midbrain as in other primates) using allometric considerations combined with a competitive-selectionist model of vocal control. However, there is more direct evidence available that pertains to the VWFA.

In their review of reading development, Schlaggar and McCandliss (2007, 484) construe the functional specialization of the VWFA in terms consistent with what Changeux and Danchin (1976) call the 'selective stabilization' of preexisting synapses. Specifically, fMRI studies conducted early in the development of reading ability show a diffuse pattern of activation throughout the extrastriate (visual) cortex. As reading skill matures, the activity becomes gradually concentrated in the region that includes the VWFA. This is consistent with a selectionist model in which the brain circuitry that most effectively mediates reading is differentially reinforced, while those circuits less appropriate are differentially weakened or eliminated. This would be an example of selection at the level of neuronal groups. Suppose, instead, one detected only a gradual strengthening of activity in the VWFA during acquisition of reading development. That alone would not implicate a selectionist model; rather, it would be consistent with a model according to which the circuitry needed for reading is created, or built, as the skill develops (see the next section on selectionist versus constructivist models of synapse formation). The available evidence lends support to selectionist models of specialization of the VWFA and therefore supports the notion that the VWFA comes to take on a novel function over the course of development.

Although little direct evidence concerning the development of the ability to play Tetris is available, the role of dopamine-mediated reward learning suggests that some *components* of this ability may also be produced by neural selection. For example, Schultz (1998, 15) presents a simple model for how dopamine neurons in the ventral tegmental area (VTA) and the substantia nigra (SN) of the brain mediate operant conditioning (instrumental learning). The activation of a certain cortical synapse (say, between neurons A and B) leads to a behavior (e.g., a successful move in Tetris). This triggers the firing of dopamine neurons in the VTA and SN, which release dopamine throughout the cortex (experienced as a pleasurable sensation). This global release of dopamine selectively strengthens the A-B synapse (as well as

others that were involved in the production of the behavior), *over connections between B and other neurons*. This leads to the reinforcement of that behavior. As a result, according to GSE, the A-B synapse would come to have the *function* of producing that behavior. Research suggests that a similar mechanism may underlie classical conditioning as well as well as instrumental conditioning (Schultz and Dickenson 2000, 18). If dopamine-mediated reward learning results in neural selection, and neural selection is sufficient for having a function, then direct proper functions are continuously coming into being in the developing brain.

Response to criticism and relation to other views

GSE can be given further elaboration and defense by showing how it can respond to criticism and by contrasting it with similar views that have been presented in the literature. First, three criticisms will be raised—the problem of liberality, the problem of distinguishing neural selection from other processes, and the relevance of selection to learning. Second, GSE will be carefully distinguished from three other attempts to define 'function' in terms of a generalized notion of selection, those of Wimsatt (1972), Millikan (1984), and Papineau (1993).

The liberality objection

The most obvious objection to GSE is that it succumbs to the same liberality problem that Boorse (1976) raised successfully against Wright's (1973) etiological theory. In Wright (1973), the function of an entity consists, roughly, in any activity that historically contributed to its own persistence, and thereby made possible the continued performance of that activity (ibid., 161). But counterexamples abound. For example, obesity contributes to a sedentary lifestyle, which in turn reinforces obesity. But obesity does not have the function of contributing to a sedentary lifestyle (Boorse 1976, 75–76). Bedau (1992, 786) imagines a stick floating down a stream that brushes against a rock and gets pinned there by the backwash it creates, and thereby perpetuates its current position, to make the same point. Nobody would want to say that the stick has the 'function' of creating a backwash.

However, such counterexamples do not threaten GSE, for the reason that none of them constitute selection processes. Although obesity contributes to a sedentary lifestyle and thereby to its own persistence, obesity is in no sense selected *over* some other trait because it contributes to a sedentary lifestyle. In other words, although obesity may contribute to its own persistence, it does not contribute to its own differential persistence (Neander 1983, 103). It simply illustrates a positive feedback loop. Similarly, the stick does not have the function of creating the backwash that pins it in place, because there is no sense in which the creation of backwash was selected over some *other* consequence because it proved more effective at pinning it in place. Because GSE incorporates the idea of a selection process, it can avoid the typical counterexamples that plagued Wright's theory.

One might think that successful counterexamples could be constructed simply by extending the Boorse-type counterexamples to describe groups that exhibit

something like differential persistence. Kingsbury (2008), for example, imagines a collection of rocks on the beach that vary with respect to their hardness, and where this variation determines how long each rock will survive before being turned into sand. This case exhibits differential persistence, yet it seems counterintuitive to say that the hardness of the rock has the function of enabling it to survive the action of the waves (ibid., 496; also see Okasha 2003, 746; Schaffner 1993, 383). However, these examples do not affect GSE because GSE explicitly restricts functions to the parts of *biological systems*. These groups are not biological systems so their parts do not have functions in this sense.

Neural selection and neural construction

One might argue that GSE faces a different sort of liberality problem in that it is too generous even within the biological context and particularly the context of neuroscience. Which types of synaptic structures result from 'neural selection' and which do not? In a sense, *all* synapse formation is 'selectionist' in that it involves the differential strengthening and weakening of synapses. If so, GSE would assign a function to all synapses, and it would not be sufficiently discriminating. As Francis Crick once remarked in a critique of 'neural Darwinism,' "almost everybody's theory [of synaptogenesis] could be called a theory of synaptic selection" (1989, 249).

In order to resolve this objection one has to carefully distinguish neural selection from other activity-dependent processes of synaptic structure formation. A neural selection process at any level to which it is applied (synapses, neurons, or groups of neurons) presupposes a population of such entities that rely for their continuation on a common 'resource.' Moreover, the uptake of this resource by one member of the population necessarily lowers the probability of uptake of that resource by other members of the same population. This is what makes it analogous to a 'competition.' This is clearly exhibited in the rat neuromuscular junction as depicted in Fig. 1 and in the phenomenon of neural cell death in which entire neurons 'compete' for a limited field of innervation. This process is not exhibited in other mechanisms of synaptic structure formation such as neural construction.

In neural construction, synaptic structures are formed by the activity-dependent production of new synapses, rather than the elimination of pre-existing ones (Purves 1994; Quartz and Sejnowski 1997). In other words, in this model, the brain triggers synapse growth on an 'as-needed' basis, thereby reducing the need for selection processes to eliminate unnecessary connections (this is the model, for example, underlying simple Hebbian networks). Often, neural construction processes serve simply to amplify and strengthen existing and frequently-used synapses. According to GSE, neural selection generates novel functions, but neural construction does not. At most, one could say of neural construction processes that they serve to *extend* and *amplify* existing functions rather than to create new ones. A proper understanding of the distinction between persistence and *differential* persistence in the context of neuroscience can provide a paradigm for thinking about differential persistence in other difficult or unusual biological cases such as clade selection, the

selection of parts of insect colonies, or the apparent evolutionary dynamics of cancer cells.

Learning and selection

A third type of criticism is that many forms of learning do not involve selection, and therefore that GSE is unable to attribute functions to many beliefs, behaviors, and dispositions that result from learning. As noted earlier, operant conditioning is one type of learning that clearly constitutes a 'selection process' in this general sense. However, there are other sorts of learning, such as learning by modeling (imitation learning), that do not seem to involve selection. Here, one must be cautious to avoid the temptation to overgeneralize the role of selection in learning.

Kingsbury (2008) is a critical response to recent attempts, such as Hull et al. 2001, to apply the selection model to learning. In particular, she notes that not all forms of learning require selection. Although this point is accurate, it is not damaging for GSE. Suppose, through imitation learning, one acquires the (rather annoying) disposition to tap one's fingers on the table. If this disposition has not been selected for, then it does not have a function. This disposition would only come to possess a function if it comes to be retained over some *other* disposition because of a rewarding consequence (e.g., by providing a source of distraction in uncomfortable situations). It is likely that many behavioral traits and beliefs do not have functions at all, but this result is plausible and consistent with common sense.

Contrast to similar views

One of the earliest attempts to generalize the concept of selection in a manner relevant to function is Wimsatt (1972), in a paper that was largely ignored in the development of the function debates. He argues 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 is clear that 'selection' does not only operate over reproducing entities; it also serves to bring about the differential reinforcement of non-reproducing entities within a system. His paradigm example is learning. At the most abstract level, he claims, all problem-solving behavior can be modeled as the outcome of a selection process, in that it involves 'blind variation' (in exploring the space of possible solutions) and 'selective retention' (in retaining the effective ones; ibid., 14).

There are two main differences between GSE and Wimsatt's view. First, Wimsatt does not restrict functions to the parts of biological systems, which GSE does. Second, he does not build selection into his *definition* of 'function' itself (Wimsatt 1972, 15–17), which is the central point of GSE. These two points are interconnected. The reason that Wimsatt does not define 'function' in terms of a selection process is that he is able to generate non-biological counterexamples to that view. For example, he points out that stars have different rates of survival (which may be a type of 'selection'), but it is counterintuitive to ascribe functions to their parts (ibid., 16). However, he could have avoided that counterexample (as does

GSE) had he defined 'function' in terms of a selection process operating only over the parts of *biological* systems.

Millikan (1984, 28) also defines 'function' in terms of a general selection process that applies both to natural selection in the evolutionary context and some types of learning. She notes that some learned behaviors possess 'functions' in the same way that biological organs do, namely, as a result of being selected for. Like Wimsatt (1972), she clearly intends the concept of selection to be a very general one. For example, if a rat's pressing a lever (rather than, e.g., pulling the lever) is associated with a reward, and as a result, the lever-pressing behavior is reproduced, it will come to possess the *function* of obtaining that reward. If a person reproduces a tool with a certain structure (rather than some other structure) because tools having said structure tend to produce a given outcome, then that tool comes to have the direct proper function of producing that outcome.[4] Her analysis of function is restricted to what she calls 'reproductively established families,' that is, groups of entities that undergo reproduction or 'copying.'

GSE differs in one crucial way from Millikan's analysis of function, namely, it does not require that a structure must have been *reproduced* in order to acquire a function. Although the neural structures underlying behavior can be *modified* or *retained* by experience, they typically are not reproduced during the individual's lifetime. Therefore, one significant limitation of Millikan's analysis of function is that, when described in their specificity, *neural structures underlying novel behaviors cannot come to possess direct functions* because they do not reproduce. Outward behavioral manifestations such as gestures can be 'reproduced' over one's lifetime but the neural structures underlying those behaviors cannot. This gives rise to the somewhat counterintuitive result that an outward behavioral trait (e.g., pushing a lever to obtain food) can have a direct proper function, but the neural structure that makes this behavior possible does not. GSE avoids this by noting that something can be selected for without being reproduced.

Papineau (1993, 44–48; also see 1987, 65–67) endorses a view of 'function' similar to that of Wimsatt (1972) and Millikan (1984). Its purpose is to recognize both evolution by natural selection as well as some forms of learning as types of 'selection processes' sufficient for generating functions. Papineau's view implicitly differs from Millikan's (and is closer to Wimsatt's) in that reproduction is not a necessary condition for having a function. A belief type or behavioral disposition can have a function by virtue of the fact that it gets 'fixed' (or reinforced) by a rewarding consequence.

The main problem with Papineau's account is that he does not carefully distinguish between the case in which a behavioral disposition (or its underlying neural substrate) is merely *reinforced* and that in which it is *differentially reinforced*

[4] This is not the primary route through which Millikan attributes functions to novel traits. Rather, she relies on a distinction between 'direct' and 'derived' proper functions to make sense of the functions of novel traits (e.g., Millikan 1984, 41–42; 1989, 288). Although this is a reasonable distinction in some contexts, it has the unattractive conclusion that one must seek out a 'direct function' for many of the novel traits that one would like to ascribe functions to, even if it is not obvious what that direct function is. For example, brain changes underlying reading ability or the ability to play Tetris may have derived proper functions, but it is not clear from which direct proper functions those functions are derived.

(that is, over some other behavior or neural substrate).[5] The difference between these two cases (mere reinforcement and differential reinforcement) is illustrated in the neurobiological context by the difference between neural construction and neural selection. Neural construction creates new synapses or reinforces existing synapses. It can lead to the strengthening of an A-B synapse even if it does not lead to the *differential* strengthening of that synapse (that is, even if it is not the case that there are connections between B and other neurons which are weakened or eliminated as a result of strengthening the A-B connection). A behavioral example can also illustrate the difference (modified from Scheffler 1958). Suppose an infant cries initially because he or she is hungry. Suppose as a result of parental solace (a type of rewarding experience) the crying is reinforced; that is, he or she cries more frequently and in the absence of hunger. The fact that the infant's crying was reinforced because, in the past, crying brought about solace, does not imply that it was *differentially* reinforced, that is, that there existed a set of variant behavior patterns culled from a pre-established repertoire (e.g., grasping, making sucking motions, etc.) that were discontinued because they failed to produce solace. Papineau's view does not distinguish between the two sorts of processes and thus is vulnerable to the liberality objection, namely, that it must extend functions to any self-perpetuating process.[6]

Conclusion and directions for further research

If GSE is right, it would show that the test cases from neuroscience which appear to contradict SE can actually be neatly assimilated to that theory. This undermines one of the long-standing criticisms of SE (described earlier) that it cannot assign functions to evolutionary novel traits in any direct way and therefore some 'forward-looking' theory must be accepted as an alternative (e.g., Schaffner 1993, 398; Walsh 1996, 558; Wouters 2003, 658; Sarkar 2005, 18).

With that criticism goes a second, related one, namely that while SE can be applied to some limited areas of evolutionary biology, it is not relevant or useful for understanding functions in other biological contexts unless evolutionary questions

[5] This ambiguity is reflected in the disparate ways he formulates the definition of 'function.' In his earlier presentation, the function of an entity is explicitly defined in terms of a 'selection process' (1987, 65). In his later presentation, the function of an entity is explicitly defined in terms of any self-perpetuating process: "*X* has the function of doing *Y* if and only if item *X* is now present *as a result* of causing *Y*" (1993, 45). In this latter view, being selected for is merely a 'paradigm' for the theory and not part of the definition.

[6] Godfrey-Smith (1992), like Papineau, explicitly includes both the products of natural selection as well as learning in an early account of biological function: "a selective basis for functional characterization is available whenever learned characters are maintained within the cognitive system because of their consequences" (ibid., 292). However, the problem with this view is that it does not specify that these learned characters must be the result of selection and not merely a generic, self-perpetuating process. Godfrey-Smith (1993, 199) rejects his earlier account of function and insists, along with Millikan, that in order for something to possess a function it must undergo reproduction, thus severing the link between evolution and learning. However, he may have avoided this admittedly unattractive conclusion (ibid.) if he had restricted functions to the parts of biological systems that undergo selection but which do not necessarily reproduce.

are being addressed (Griffiths 2006, 3). On the contrary, SE functions and CR functions would appear to embody different types of information about traits that would be valuable in any biological context. While SE embodies historical information (including developmental as well as evolutionary), the causal role theory embodies information about how a part currently contributes to a phenomenon of research interest.[7]

There are two key advantages to accepting GSE as an account of biological function for neuroscience. As has been shown, selection processes can plausibly be used to assign functions to all three of the 'test cases' in neuroscience and they do so in a way that preserves the elegance of SE. What is particularly elegant is that in all three cases, the same general process (selection) in invoked, although it operates at different levels and in different domains. Secondly, etiological accounts of function in general can easily and naturally preserve widespread intuitions about function statements, specifically, the intuitions that function statements are *explanatory* and *normative*. Neither intuition sits comfortably with CR. Given that SE preserves widespread intuitions about function statements, and can be applied in biologically reasonable ways, there are strong *prima facie* reasons for regarding it as a complementary and informative approach to understanding functions in the context of neuroscience.

While selection processes have been shown to take place in the context of evolutionary biology, neuroscience, immunology, and some forms of learning, it remains to be seen whether selection processes take place in other realms as well. One problem-context is ecology. One of the peculiar features of the concept of function in ecology is that ecologists assign functions to populations *as such*. This is almost incoherent on the standard selected effects theory, which typically only assigns functions to specific biological traits within individual organisms (Maclaurin and Sterelny 2008, 114). However, is there any sense in which a population or even a non-biotic ecosystem feature such as a soil type is *differentially retained* within an ecosystem by virtue of its historical role in contributing to the persistence of the ecosystem? Bouchard (forthcoming) argues that for the context of ecology, function should be construed in terms of the contribution an ecosystem component makes to the differential persistence of that ecosystem itself (that is, over other ecosystems; ibid.). Thus he acknowledges that something like differential persistence can be applied to entire ecosystems. However, his view differs in two ways from GSE. First, unlike GSE, Bouchard embraces a forward-looking approach to functions—he eschews history and defines function in terms of the propensity of a given ecosystem component to contribute to the differential persistence of the ecosystem. Secondly, Bouchard construes the entire ecosystem as a unit of selection, and then identifies the functions of ecosystem components in terms of their contribution to the selection of the ecosystem as a whole (over other ecosystems). However, GSE would identify the function of an ecosystem component, such as a population, in terms of the activity that historically contributed to its own differential persistence over other populations (perhaps via its contribution to the persistence of the ecosystem as a whole). It need

[7] Brandon (forthcoming) draws a similar conclusion by pointing to the valuable co-existence of historical and ahistorical concepts in many scientific disciplines.

not countenance the idea that ecosystems themselves can differentially persist. This would represent a valuable and potentially fruitful extension of the etiological theory into ecology. However, this prospect awaits more detailed evaluation.

Acknowledgments The author wishes to express his gratitude to Carl Craver, Gualtiero Piccinini, Anya Plutynski, and the University of Utah philosophy of biology lab group for comments and criticism on an earlier draft.

References

Allen C, Bekoff M (1995) Biological function, adaptation, and natural design. Philos Sci 62:609–622
Amundson R, Lauder GV (1994) Function without purpose: the uses of causal role function in evolutionary biology. Biol Philos 9:443–469
Antonini A, Stryker MP (1993) Rapid remodeling of axonal arbors in the visual cortex. Science 260:1819–1821
Bedau M (1992) Where's the good in teleology? Philos Phenomenol Res 52:781–805
Bigelow J, Pargetter R (1987) Functions. J Philos 84:181–196
Blair RJ (2003) Neurological basis of psychopathy. Br J Psychiatry 182:5–7
Boorse C (1976) Wright on functions. Philos Rev 85:70–86
Bouchard F (2008) Causal processes, fitness, and the differential persistence of lineages. Philos Sci 75:560–570
Bouchard F (forthcoming) How ecosystem evolution strengthens the case for functional pluralism. In: Huneman P (ed) Functions: selection and mechanisms. Synthese Library, Boston, pp 56–71
Brandon RN (1990) Adaptation and environment. Princeton University Press, Princeton
Brandon RN (forthcoming) A general case for functional pluralism. In: Huneman P (ed) Functions: selection and mechanisms. Synthese Library, Boston, pp 72–78
Brown MC, Jansen JKS, Van Essen D (1976) Polyneural innervation of skeletal muscle in new-born rats and its elimination during maturation. J Physiol 261:387–422
Burnet FM (1959) The clonal selection theory of acquired immunity. Cambridge University Press, Cambridge
Changeux J-P, Danchin A (1976) Selective stabilization of developing synapses as a mechanism for the specification of neuronal networks. Nat 264:705–711
Craver C (2001) Role functions, mechanisms, and hierarchy. Philos Sci 68:53–74
Craver C (forthcoming) Functions and mechanisms: a perspectivalist view. In: Huneman P (ed) Functions: selection and mechanisms. Synthese Library, Boston, pp 199–220
Crick F (1989) Neural Edelmanism. Trends Neurosci 12:240–248
Cziko G (1995) Without miracles: universal selection theory and the second Darwinian revolution. MIT Press, Cambridge
Darden L, Cain JA (1989) Selection type theories. Philos Sci 56:106–129
Davies PS (2001) Norms of nature: naturalism and the nature of functions. MIT Press, Cambridge
Deacon TW (1989) The neural circuitry underlying primate calls and human language. Hum Evol 4:367–401
Edelman G (1978) Group selection and phasic reentrant signaling: a theory of higher brain function. In: Edelman GE, Mountcastle VB (eds) The Mindful brain: cortical organization and the group-selective theory of higher brain function. MIT Press, Cambridge, pp 51–100
Edelman G (1987) Neural Darwinism: the theory of neuronal group selection. Basic Books, New York
Garson J (2008) Function and teleology. In: Sarkar S, Plutynski A (eds) A companion to the philosophy of biology. Blackwell, Malden, MA, pp 525–549
Garson J (2010) Schizophrenia and the dysfunctional brain. J Cog Sci 11:215–246
Godfrey-Smith P (1992) Indication and adaptation. Synthese 92:283–312
Godfrey-Smith P (1993) Functions: consensus without unity. Pac Philos Q 74:196–208
Godfrey-Smith P (1994) A modern history theory of functions. Noûs 28:344–362
Godfrey-Smith P (2007) Conditions for evolution by natural selection. J Philos 104:489–516
Gould SJ, Lewontin RC (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc B 205(1161):581–598

Griffiths PE (1993) Functional analysis and proper function. Br J Philos Sci 44:409–422
Griffiths PE (2006) Function, homology, and character individuation. Philos Sci 73:1–25
Haier RJ, Karama S, Leyba L, Jung RE (2009) MRI assessment of cortical thickness and functional activity changes in adolescent girls following three months of practice on a visual-spatial task. BMC Res Notes 2:174–180
Hardcastle VG (1999) Understanding functions: a pragmatic approach. In: Hardcastle VG (ed) Where biology meets psychology: philosophical essays. MIT Press, Cambridge, pp 27–43
Hull DL, Langman RE, Glenn SS (2001) A general account of selection: biology, immunology and behavior. Behav Brain Sci 24:511–527
Kingsbury J (2008) Learning and selection. Biol Philos 23:493–507
Kitcher P (1993) Function and design. Midwest Stud Philos 18:379–397
Lewens T (2007) Functions. In: Matthen M, Stevens C, Gabbay DM, Thagard P, Woods J (eds) Philosophy of biology. Elsevier, Amsterdam, pp 525–547
Mace CA (1949 [1935]) Mechanical and teleological causation. In: Feigl H, Sellars W (eds) Readings in philosophical analysis. Appleton-Century-Crofts, Inc, New York, pp 534–539
MacLaurin J, Sterelny K (2008) What is biodiversity? University of Chicago Press, Chicago
Mathews TJ, Ventura SJ (1997) Birth and fertility rates by educational attainment: United States, 1994. Mon Vital Statistics Rep 45(10):1–19
Matteo M, Saborido C, Moreno A (2009) An organizational account of biological functions. British J Philos Sci 60:813–841
McDowell JJ (2009) Behavioral and neural Darwinism: selectionist function and mechanism in adaptive behavior dynamics. Behav Process. doi:10.1016/j.beproc.2009.11.011
McLaughlin P (2001) What functions explain: Functional explanation and self-reproducing systems. Cambridge University Press, Cambridge
Millikan RG (1984) Language, thought, and other biological categories. MIT Press, Cambridge
Millikan RG (1989) In defense of proper functions. Philos Sci 56:288–302
Mitchell SD (1995) Function, fitness, and disposition. Biol Philos 10:39–54
Neander K (1983) Abnormal psychobiology. Dissertation, La Trobe
Neander K (1991) Functions as selected effects: the conceptual analyst's defense. Philos Sci 58:168–184
Okasha S (2003) Does the concept of "clade selection" make sense? Philos Sci 70:739–751
Papineau D (1987) Reality and representation. Blackwell, New York
Papineau D (1993) Philosophical naturalism. Blackwell, Oxford
Petersen SE, Fox PT, Snyder AZ, Raichle ME (1990) Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. Sci 249:1041–1044
Pettmann C, Henderson CE (1998) Neuronal cell death. Neuron 20:647–653
Prior EW (1985) What is wrong with etiological accounts of biological function? Pac Philos Q 66:310–328
Purves D (1994) Neural activity and the growth of the brain. Cambridge University Press, Cambridge
Purves D, Lichtman JW (1980) Elimination of synapses in the developing nervous system. Sci 210:153–157
Quartz SR, Sejnowski TJ (1997) The neural basis of cognitive development: a constructivist manifesto. Behav Brain Sci 20:537–596
Rajewsky K (1996) Clonal selection and learning in the antibody system. Nat 381:751–758
Robins SK, Craver CF (2009) Biological clocks: explaining with models of mechanisms. In: Bickle J (ed) The oxford handbook of philosophy and neuroscience. Oxford University Press, Oxford, pp 41–67
Ruse ME (1971) Functional statements in biology. Philos Sci 38:87–95
Sarkar S (2005) Molecular models of life: Philosophical papers on molecular biology. MIT Press, Cambridge, MA
Schaffner K (1993) Discovery and explanation in biology and medicine. University of Chicago Press, Chicago
Scheffler I (1958) Thoughts on teleology. Br J Philos Sci 9:265–284
Schlaggar BL, McCandliss BD (2007) Development of neural systems for reading. Ann Rev Neurosci 30:475–503
Schultz W (1998) Predictive reward signal of dopamine neurons. J Neurophysiol 80:1–27
Schultz W, Dickenson A (2000) Neuronal coding of prediction errors. Ann Rev Neurosci 23:473–500
Schwartz PH (1999) Proper function and recent selection. Philos Sci 66:S210–S222
Sober E (1984) The nature of selection. MIT Press, Cambridge
Walsh DM (1996) Fitness and function. British J Philos Sci 47:553–574

Walsh DM, Ariew A (1996) A taxonomy of functions. Can J Philos 26:493–514
Wiesel TN, Hubel DH (1963) Single-cell responses in striate cortex of kittens deprived of vision in one eye. J Neurophysiol 26:1003–1017
Wimsatt WC (1972) Teleology and the logical structure of function statements. Stud Hist Philos Sci 3:1–80
Wong ROL, Lichtman JW (2002) Synapse elimination. In: Squire LR, Bloom FE, McConnell SK, Roberts JL, Spitzer NC, Zigmond MJ (eds) Fundamental neuroscience, 2nd edn. Academic Press, Amsterdam, pp 533–554
Wouters A (2003) Four notions of biological function. Stud Hist Philos Biol Biomed Sci 34:633–668
Wright L (1973) Functions. Philos Rev 82:139–168