Published in final edited form as:

Int J Law Psychiatry. 2011 ; 34(1): 20–29. doi:10.1016/j.ijlp.2010.11.004.

Who's flying the plane: Serotonin levels, aggression and free will

Allan Siegel^{a,*} and John Douard^b

^aDepartments of Neurology & Neuroscience and Psychiatry, University of Medicine & Dentistry of NJ—NJ Medical School, 185 South Orange Avenue, Newark, NJ 07103, USA

^bRutgers University, Lecturer, Department of Philosophy, University of Medicine & Dentistry of NJ, School of Public Health, Adjunct Professor, Criminal Defense Appellate Attorney

Abstract

The present paper addresses the philosophical problem raised by current causal neurochemical models of impulsive violence and aggression: to what extent can we hold violent criminal offenders responsible for their conduct if that conduct is the result of deterministic biochemical processes in the brain. This question is currently receiving a great deal of attention among neuroscientists, legal scholars and philosophers. We examine our current knowledge of neuroscience to assess the possible roles of deterministic factors which induce impulsive aggression, and the extent to which this behavior can be controlled by neural conditioning mechanisms. Neural conditioning mechanisms, we suggest, may underlie what we consider the basis of responsible (though not necessarily moral) behavior: the capacity to give and take reasons. The models we first examine are based in part upon the role played by the neurotransmitter, serotonin, in the regulation of violence and aggression. Collectively, these results would appear to argue in favor of the view that low brain serotonin levels induce impulsive aggression which overrides mechanisms related to rational decision making processes. We next present an account of responsibility as based on the capacity to exercise a certain kind of reason-responsive control over one's conduct. The problem with such accounts of responsibility, however, is that they fail to specify a neurobiological realization of such mechanisms of control. We present a neurobiological, and weakly determinist, framework for understanding how persons can exercise guidance control over their conduct. This framework is based upon classical conditioning of neurons in the prefrontal cortex that allow for a decision making mechanism that provides for prefrontal cortical control of the sites in the brain which express aggressive behavior that include the hypothalamus and midbrain periaqueductal gray. The authors support the view that, in many circumstances, neural conditioning mechanisms provide the basis for the control of human aggression in spite of the presence of brain serotonin levels that might otherwise favor the expression of impulsive aggressive behavior. Indeed if those neural conditioning mechanisms underlie the human capacity to exercise control, they may be the neural realization of reason-responsiveness generally.

Keywords

aggressive behavior; determinism; free will; hypothalamus; midbrain periaqueductal gray; serotonin

© 2010 Elsevier Ltd. All rights reserved.

*Corresponding author. Department of Neurology & Neuroscience, New Jersey Medical School, MSB-H-592, 185 South Orange Avenue, Newark, NJ 07103, USA. Tel.: +1 973 972 4471; fax: +1 973 972 3291. Siegel@UMDNJ.edu (A. Siegel).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

One of the most formidable tasks that confront both professionals and laymen in the legal and health care professions is the challenge of evaluating and determining the degree of culpability that may be ascribed to persons who commit criminal offenses. Culpability is ascribed on the basis of conditions that permit us to hold a person responsible for blameworthy actions. Of particular concern are aggressive actions that violate the rights of others. A classic example is one in which an individual impulsively kills another person. How do we determine the extent to which the individual displayed moral culpability for her act? A state's punitive response to such an act requires finding that the offender acted while in a certain mental state, or *mens rea*. In the United States, criminal justice systems do not recognize a special class of violent behavior that can be characterized as "impulsive." In the Model Penal Code, which has influenced the criminal code of most states, culpability is defined in terms of an act's being "purposeful," "knowing," "reckless," or "negligent." Because each of these mental states involves some measure of control, culpability is usually linked to our capacity to control our behavior.

Recent advances in cognitive neuroscience challenge the assumption that we can be held responsible for our actions, including actions for which the criminal justice system holds us criminally culpable. If it can be shown, for example, that a dominant causal factor of a person's committing an impulsive violent crime is a neurochemical process in the brain, is the offender's action sufficiently under her control to warrant attributing the mental state(s) required to render a guilty verdict? Traditionally, this practical question invokes the more general philosophical question whether human action is the outcome of the exercise of "free will," or is "determined" by factors over which we have no control. The current version of this philosophical question focuses on neurochemical events in the brain in conjunction with environmental events. The following discussion considers initially how neurochemical models, which might support a deterministic explanation of violent behavior, can explain violent impulsive behavior. We then argue that neural determinism is compatible with a notion of control that implicates the human capacity to give and take reasons for conduct. This is an ancient notion of responsibility that is anchored in practical reason rather than in a metaphysical notion of "free will." We conclude with an admittedly speculative sketch of a possible neural mechanism that might explain, within a deterministic framework, the causal role of brain events in the exercise of intentional control over conduct.

In considering how neural mechanisms may help to provide a better understanding of violent impulsive aggression, the focus of the present discussion is upon the role of the neurotransmitter, serotonin. This is not to say that other neurotransmitters and neuromodulators do not have an important role to play in the expression and control of violent impulsive aggression. But as described below, the evidence in support of a potent serotonin mechanism appears to be overwhelming and for this reason, the discussion is aimed at providing an understanding of the actions of this neurotransmitter.

2. Serotonin (5-HT) synthesis and removal

Brain serotonin levels have been found to be related to certain types of aggressive behavior. Specifically, a decrease in serotonin levels is inversely proportional to aggression. It follows that altering serotonin levels will affect behavior in ways that may be relevant to the question whether aggressive behavior is caused by processes over which we have little or no conscious control. Because we want to tie our analysis to the details of neuroscience, we shall first review some of the studies that support the hypothesis that serotonin plays a causal role in modulating aggressive behavior. We assume that if studies utilizing several different

methodologies for regulating serotonin levels all support the hypothesis, the hypothesis has a high degree of probability and predictive power.

Serotonin does not cross the blood-brain barrier. Therefore, brain serotonin results from the synthesis in selective brain cells situated in the raphe nuclei of the medulla, pons, and midbrain. The substrate for serotonin formation is dietary tryptophan, which enters the brain through an uptake process. In brainstem raphe neurons, it is hydroxylated to form 5-hydroxytryptophan by tryptophan hydroxylase. It is then decarboxylated to form serotonin by aromatic-L-amino acid decarboxylase and stored in vesicles and ultimately released onto neurons with which it makes synaptic contact. The mechanism for removal of serotonin from the synaptic cleft is by reuptake and metabolism. Serotonin is deaminated to 5-hydroxyindoleacetaldehyde by monoamine oxidase and is then oxidized to form 5-hydroxyindolacetic acid by aldehyde dehydrogenase and ultimately secreted through the urine.

From the description of serotonin synthesis and removal described above, it should be readily seen that studies could be designed to determine the effects of serotonin and violent behavior through modification of brain levels of serotonin by: (1) altering dietary levels of serotonin, (2) controlling enzyme synthesis, (3) serotonin receptor activation or blockade, (4) inhibiting reuptake of serotonin, and (5) genetic manipulation of serotonin and its receptors. In addition, the effects of serotonin upon violent behavior could also be discerned by correlating brain levels of serotonin in individuals with their propensity to express aggression and rage (or other forms of violent) behavior. The following discussion summarizes some of the key data that utilized one or more of the methodologies related to manipulation of serotonin levels in the brain as described above. More detailed descriptions summarizing larger bodies of data regarding the role of serotonin in aggression and impulsive behavior can be found in review articles by Siegel, Bhatt, Bhatt and Zalcman (2007) and Krakowski (2003).

3. Studies testing the role of the serotonin system in aggression, violence and impulsive behavior

3.1. Evidence of a causal relationship--1: Effects of acute tryptophan depletion

Prior to discussion of human studies upon aggression, one point of caution ought to be addressed concerning studies in which either a psychometric analysis or quantification of personality traits is applied. Such scales can either be nominal, ordinal, or linear in nature. Thus, the significance of the data summarized may differ from study to study and therefore, comparisons of statistical analyses among the studies as well as the overall conclusions drawn from data reported should be viewed with a degree of caution.

One way to test the hypothesis that serotonin modulates aggression is to regulate the intake of dietary tryptophan, the amino acid precursor essential for the synthesis of serotonin and to then determine the effects of such manipulation upon aggressive behavior. Several studies are reported here. In the first, Williams, Shoaf, Hommer, Rawlings & Linnoila (1999) attempted to establish the fact that dietary depletion of tryptophan can, in fact, reduce serotonin levels in the brain. Six normal male and female adults who were in good health were placed on a low tryptophan diet that included a drink that previously had been shown to deplete plasma tryptophan (Delgado, Charney, Price, Aghajanian, Landis & Heninger, 1990). Estimates of brain serotonin levels were determined by lumbar puncture in which the serotonin metabolite in cerebrospinal fluid (CSF), 5-hydroxy-indolacetic acid (5-HIAA), was measured at different time points following administration of the amino acid depleting liquid. It is well established that CSF 5-HIAA provides a very close estimate of brain serotonin levels (Brunton, Lazo, & Parker, 2006; Echizen & Freed, 2006; Raab, A., 1970).

The results clearly demonstrated that delivery of the amino acid depleting liquid resulted in significant lowering of plasma and CSF tryptophan as well as 5-HIAA in the CSF. Thus, this study provided the empirical basis by which dietary alterations in levels of brain serotonin could be estimated by the measurement of 5-HIAA in the CSF and correlated with changes in aggressive behavior as was shown in the following study.

Using paid volunteers, Moeller, Dougherty, Swann, Collins, Davis & Cherek (1996) utilized a point-subtraction task in which the subject had the choice of selecting a non-aggressive response (i.e., accumulating points that could be exchanged for money), or an aggressive response (i.e., which subtracts points from a fictitious individual). The behavioral paradigm enabled the experimenter to provoke a subject by periodically subtracting points from him, which the subject believed was carried out by the fictitious person with whom he was paired. The extent of point subtraction was measured both prior to and following administration of a tryptophan free diet for 24 hours. The results demonstrated a significant increase in aggression as early as 5 hours following ingestion of the low tryptophan diet relative to baseline when no dietary changes were instituted. Likewise, there was a significant decrease in plasma tryptophan 5 hours following ingestion of the low tryptophan diet.

In one other study noted here, a similar aggression paradigm was applied to the study of aggressive behavior in women (Marsh, Dougherty, Moeller, Swann and Spiga, 2002). In this study, similar findings were reported to those described above; namely, that women subjected to tryptophan depletion displayed elevated levels of aggression scores. Moreover, when dietary quantities of tryptophan were increased, there was a decrease in aggression scores. Thus, these studies provide support for the hypothesis that dietary induced reduction in brain serotonin is associated with increases in aggressive behavior.

3.2. Evidence of a causal relationship – 2: Pharmacological manipulation of brain serotonin levels

A number of studies have utilized pharmacological tools as a means of manipulating brain serotonin levels in order to test the role of brain serotonin in aggressive behavior. One method applied in several studies conducted by Coccaro and colleagues. Typically, fluoxetine (Prozac) was administered to 40 non-major depressed, non-bipolar or schizophrenic personality-disordered individuals who at the time of the study expressed impulsive, aggressive behavior and irritability. The results indicated that fluoxetine, but not placebo administration resulted in sustained reductions in scores on the Irritability and Aggression subscales of the Overt Aggression Scale (Coccaro and Kavoussi, 1997). It is interesting to note that similar findings were reported in a study involving Vervet monkeys. Fairbanks, Melega, Jorgensen, Kaplan and McGuire (2001) observed that initially, aggressive interactions were inversely correlated with levels of 5-HIAA in the CSF. In addition, daily treatment with fluoxetine was found to significantly lower impulsivity in treated monkeys relative to control subjects.

A related approach has been employed with the use of the indirect serotonin agonist, fenfluramine (which enhances the availability of serotonin at the postsynaptic region). The rationale for this approach is based upon the fact that the release of serotonin at central synapses induces the release of pituitary hormones such as prolactin, which can be easily measured in response to fenfluramine challenge. Dolan and Anderson (2003) studied personality disordered offenders and reported an inverse relationship between the prolactin response to fenfluramine challenge and measures of antisocial conduct, a finding parallel to those reported earlier by Coccaro, Kavoussi, Cooper and Hauger (1997) and Coccaro, Kavoussi, Trestman, Garbriel, Cooper and Siever (1997). In support of these findings, Cleare and Bond (1997) further reported an inverse relationship between fenfluramine and

cortisol responses and scores on the Buss-Durkee Hostility Inventory aggression and total inventory scores in healthy and normal adult males.

3.2. Evidence of a relationship between brain serotonin and aggression as suggested from correlational studies

Another approach to the study of the relationship between serotonin and aggression has been to attempt to correlate brain serotonin levels with the propensity to express aggressive behavior. A number of studies have provided data whose findings are basically similar.

An early study by Asberg, Traskman & Thoren (1976) provided evidence that CSF 5-HIAA levels can be used as a predictor of suicide. More recently, studies by Lidberg and colleagues showed that reduced serotonin levels in platelets (Lidberg & Daderman, 1997), or low levels of 5-HIAA in cerebrospinal fluid are related to low impulse control (Lidberg, Belfrage, Bertisson, Evenden, & Asberg, 2000), or to a relapse in violent crime (Daderman & Lidberg, 2002).

In considering one study in greater detail, it was shown that CSF-5-HIAA levels were significantly lower in impulsive violent attempters at suicide than non-violent attempters (Cremniter, Jamian, Kollenbach, Alvarez, Lecrubier, Gilton, Jullien, Lesieur, Nonnet and Spreux-Varoquaux, 1999). When a related study was conducted utilizing violent subjects who have had lifetime violence and hostility ratings in comparison to non-violent subjects, it was reported that, likewise, the violent subjects had lower CSF-5-HIAA concentrations than nonviolent subjects (Hibbeln, Umhau, Linnoila, George, Ragan, Shoaf, Vaughan, Rawlings, and Salem Jr, 1998). While there is sometimes a linkage between suicidal tendencies and violent behavior in the studies conducted, a further study specifically selected 64 patients for analysis who had no past history of suicide behavior (Stanley, Molcho, Stanley, Winchel, Gamaroff, Parsons, and Mann, 2000). These authors observed that, when aggressive patients were separated from those who were non-aggressive, it was observed that aggressive patients had significantly lower concentrations of CSF-5-HIAA than non-aggressive patients (although correlations with self reports of impulsivity and hostility were not apparent in this study). It is of further interest to point out that similar findings were observed in a study of rhesus monkeys (Mehlman, Higley, Faucher, Lilly, Taub, Vickers, Suomi and Linnoila, 1994). In this study, monkeys were fitted with radio transmitters for rapid identification of their locations in a natural habitat and could then be observed for the extent to which aggressive and submissive behaviors occur within their colonies. The results indicated that male rhesus monkeys with low CSF-5-HIAA concentrations were at higher risk for expressing violent, aggressive responses and loss of impulse control as indicated by greater risk taking behavior. Thus, this study, coupled with those described above, provide additional evidence of the linkage between low levels of brain serotonin and heightened levels of aggression and impulsive behavior in both humans and non-human primates.

3.3. Evidence of a relationship between serotonin and aggression as determined by genetic studies

One approach to the study of genetics of aggression has been to utilize genetic engineering to test the role of serotonin receptors in the regulation of aggressive behavior. In a study conducted in mice, Sandou, Amara, Dierich, LeMeur, Ramboz, Segu, Buhot and Hen (1994) showed that mice bred specifically that lacked the serotonin 5-HT_{1B} receptor were found to be significantly more aggressive upon provocation than control animals. At the human level, several studies are described here.

A totally separate and distinct approach was adapted by Constantino, Morris and Murphy (1997), who attempted to compare 5-HIAA levels in newborns of personality disordered

parents relative to normal parents. In this study, assays of 5-HIAA were taken from the CSF from 193 neurologically normal newborn infants and family psychiatric histories were taken of the first- and second-degree relatives of the newborns. It was observed that 5-HIAA levels of newborns whose parents had family histories of antisocial personality disorder were lower than in newborns whose parents had normal family histories. Several opposing conclusions are suggested from these findings. One view suggests that personality-disordered parents may have lower than average levels of brain serotonin that are genetically transmitted to their offspring. Such a view would further imply that, because these newborns may have low serotonin levels, they may be more prone to the expression of aggressive behavior at a later age than those whose parents are viewed as normal and whose brain serotonin levels are somewhat higher. An alternative interpretation of the data might suggest that the lower serotonin levels of the infants were generated, not on a genetic basis, but as a result of a different and perhaps somewhat more hostile environment that may exist in the homes of personality-disordered parents which is somehow sensed by the infants of these parents.

Another approach to the study of genetic factors related to serotonin and aggressive behavior involves an analysis of the monoamine oxidase gene. In one study, it was concluded that maltreatment and mental health problems are associated with low vs. high monoamine oxidase activity (Kim-Cohen, Caspi, Taylor, Williams, Newcombe, Craig, & Moffitt, 2006). In a separate study, an analysis was conducted of the alleles of the monoamine oxidase gene in normal and aggressive individuals (Reif, Rösler, Freitag, Schneider, Eugen, Kissling, Wenzler, Jacob, Retz-Junginger, Lesch and Retz, 2007). This analysis was selected in part because it is well known that monoamines, including serotonin, are degraded by monoamine oxidase. The basic study involved characterization of individuals on a composite measure of aggression and associated variables with respect to monoamine oxidase gene alleles that these individuals carry. The results indicated a general pattern in which a larger majority of people associated with violent behavior carried the low-activity, short monoamine oxidase allele. These authors further observed that this relationship occurs principally as an interaction between high adversity in childhood and violence in adulthood where the short promoter alleles were present. Thus, this study, as well as others described above, indicates that genetic factors are important in the expression of aggressive behavior, but such effects are also linked to environmental conditions during behavioral development.

3.4. Effects of a drug of abuse on serotonin levels and aggressive behavior – the role of alcohol

In considering the role of serotonin in the regulation of aggressive behavior, it has been of interest to investigators to question whether drugs of abuse, which are known to potentiate aggressive behavior, manifest their effects through a serotonin mechanism. In particular, alcohol constitutes just such a drug of abuse and has been the subject of many studies in both animals and humans. Several studies are summarized here. One study was designed to determine whether or not alcohol consumption can affect serotonin metabolism (Badawy, 1998). In this study, a normal population of non-aggressive individuals was administered 2–2.5 pints of normal strength beer (approximately 1 liter) which resulted in an elevation of blood alcohol levels after 1 hr to 75–78 mg/dl, which is just below the legal limit of 80 mg/dl for driving in England. Serum tryptophan concentrations were measured and revealed that after 2 hr, there were dramatic decreases in serum tryptophan concentrations in these subjects that began to recover at 3 hr, post-alcohol administration. From these data, it would be implied that the levels of brain serotonin would decrease in a parallel fashion, as described above by Williams, Shoaf, Hommer, Rawlings and Linnoila (1999). Thus, a lowering of serotonin levels by alcohol consumption could easily serve as the mechanism for the presence of aggressive behavior often observed by some individuals soon after they

have consumed alcohol. Consistent with this view, a study by Pihl, Young, Harden, Plotnick, Chamberlain and Ervin (1995) and summarized in Pihl and Lemarquand (1998) demonstrated that, when normal healthy volunteers were administered acute tryptophan depletion coupled with alcohol, they displayed heightened levels of aggressive behavior as indicated by their delivery of higher and longer shock intensities to (non-existent) opponents relative to those who either received augmented tryptophan mixtures or who were not administered alcohol. Collectively, these data point to the importance of alcohol in potentiating aggressive behavior via a serotonergic mechanism and its significance as a serious risk factor for individuals with aggressive tendencies.

4. The neural mechanism for the expression of affective (hostile) aggressive behavior

In attempting to understand the sites and regions of the brain where serotonin may interact with serotonin receptors to modulate neuronal activity that directly affects aggressive behavior, it is instructive to first identify the sites and regions where such interactions may likely occur which are associated with the type of aggression examined. Specifically, the form of aggression most frequently studied in both animal and human studies, including those described earlier in this paper, is referred to as affective defense or defensive rage, which comprises the primary segments of the animal literature involving felines and hostile aggression in the human literature. It should also be noted that the significance of the use of animal research with respect to its application to our understanding of the neural bases of human aggression is based upon two critical factors: (1) that the types of aggression observed in animals significantly overlap with those described in humans; and (2) likewise, the regions of the brain critical for the expression and modulation of aggression in animals parallel those in humans.

Major components of the neural circuitry, including the primary regions of the brain associated with this form of aggression, are now well established (Fuchs, Edinger and Siegel, 1985; Shaikh, Barrett and Siegel, 1987; Siegel, Roeling, Gregg and Kruk, 1999; Siegel, 2005; Siegel, Bhatt, Bhatt and Zalzman, 2007; Siegel and Victoroff, 2009). In brief, the major sites in the brain where integration of this form of aggression takes place include the medial hypothalamus and the dorsal aspect of the midbrain periaqueductal gray matter (PAG). Essentially, glutamatergic neurons mediating this behavior which arise principally from the anterior third of the medial hypothalamus project directly to the dorsal PAG, exciting neurons in this region by acting through NMDA receptors. The PAG, in turn, projects to somatomotor and autonomic regions of the lower brainstem which include motor nuclei of cranial nerves V, VII and possibly IX and X for the vocalization components of affective aggression and to the solitary nucleus, and indirectly to the ventrolateral medulla (which projects directly to sympathetic nuclei of the spinal cord) for regulation of the autonomic components of this response. Damage to components of this output system severely disrupts the capacity of the organism to express rage behavior. Of particular relevance to the present discussion, it is now known that serotonin receptors are abundantly present in the medial hypothalamus and PAG and that both regions receive significant serotonergic inputs from raphe nuclei (Shaikh, De Lanerolle, and Siegel, 1997; Hassanain, Bhatt and Siegel, 2003). Moreover, recent studies have revealed that activation of serotonin receptors in these regions potently modulate affective forms of aggression in the cat (Shaikh, De Lanerolle, and Siegel, 1997; Hassanain, Bhatt and Siegel, 2003; Bhatt, Bhatt, Zalzman and Siegel, 2009). Therefore, it is likely that modulation of aggression by adjustment of levels of brain serotonin is manifest through its actions upon neurons in the medial hypothalamus and PAG. It should also be noted that it is possible that serotonin's effects upon aggressive behavior may occur by its actions on limbic structures such as the prefrontal cortex (Frokjaer, Mortensen, Nielsen, Haugbol, Pinborg, Adams, Svarer,

Hasselbalch, Holm, Paulson, and Knudsen, 2008), which is known to powerfully modulate aggression and rage behavior (Siegel, 2005). While serotonin 5-HT_{1A} suppresses and 5-HT₂ receptors facilitate affective aggression (Shaikh, De Lanerolle, and Siegel, 1997; Hassanain, Bhatt and Siegel, 2003), it may be concluded that the predominant effect acts through 5-HT_{1A} receptors to suppress impulsive aggression. Thus, it is this mechanism that governs serotonergic control of affective aggression, and when this mechanism is diminished or suppressed, the likelihood of release of an affective, impulsive aggressive response is increased.

5. Conclusions suggested from our knowledge of the role of serotonin as an important component of the neural mechanisms of aggression: implications for a deterministic view underlying the expression of impulsive affective aggression

From the discussion above, it is reasonable to conclude that the propensity to express impulsive affective aggressive behavior is determined by environmental and genetic factors that modify brain serotonin (and presumably other brain neurotransmitters and neurochemical compounds) that interact with the neuronal elements comprising the basic nuclear groups that provide the substrate for the expression of this behavior. According to this view, the likelihood of an individual committing a violent crime could be related to the extent to which the person presents with significant variations in brain neurochemistry. One might argue that this condition with respect to aggressive tendencies would be analogous to that where an individual suffering from a hypokinetic movement disorder such as Parkinson disease has little control over bodily movements (such as a pill rolling tremor) due to her loss of dopamine neurons in the substantia nigra. Thus, with respect to the aggressive individual, we may conclude that, on the basis of studies described earlier in this paper, she had little control over her behavior because of a neurochemical imbalance.

This is the conclusion that is often drawn by neuroscientists interested in the apparent conflict between determinism and “free will.” Thus, Joshua Green, trained in both philosophy and cognitive neuroscience, argues that the advances in neuroscience show that determinism is true, and *therefore* we need to abandon the metaphysics of free will, and the folk-morality of responsibility and culpability grounded in that metaphysics (Green & Cohen, 2004). While we hold no brief for archaic metaphysics, including a Cartesian notion of a will that regulates conduct free from the constraints of physical processes, it is also clear to us that no direct conclusions about responsibility and culpability can be drawn from what we know about neurobiology. To clarify our position, some conceptual analysis is required.

6. The philosophical context: compatibilism vs. determinism¹

Below, we defend a neuroscience-informed version of *compatibilism*: the view that human behavior is part of the causal structure of the world, and can be explained in terms of causal laws or powers, but that human beings are capable of exercising sufficient control over their conduct that we can hold them morally responsible. In this limited sense, free will and a weak form of determinism are mutually compatible. A strong form of determinism, which we consider largely ideological, holds that human behavior is no more under the control of the person as agent than the behavior of billiard balls set in motion.

¹A third theory, often called “libertarianism,” holds that determinism is false, because human behavior is not caused by antecedent physical events. We do not consider this view in any detail here, except to note that if human behavior is uncaused, that is, indeterminate, then it is not under the control of an agent who can be held responsible for her behavior (Dupre, 1996).

Very few, if any, philosophers or neuroscientists today believe that human conduct occurs free from all antecedent causes or are the outcome of non-physical mental events. Our argument here is a species of compatibilism because we believe the normative structure of the social world, including scientific practice, is an outcome of the causal capacities of certain types of organized system: systems to which we generally attribute reasons, desires, and beliefs. What has been called “folk psychology,” not without sarcasm, is an ineliminable feature of psychological explanation, but nonetheless compatible with neuroscience. However, we also consider folk-psychological capacities to be biological and causally related to neural events. There is no responsible ghost in the non-responsible machine.

While it is beyond the scope of this paper to present, even briefly, the history of compatibilism, we shall lay out the common features of many of the arguments for compatibilism. To anticipate our broad conclusion, we live in what philosopher of physics Nancy Cartwright (1999) calls a “dappled world,” a world in which the most important discoveries in science are not “laws of nature,” but discoveries about the “natures of things,” including us. Scientific knowledge is fundamentally knowledge of the capacities of diverse kinds of thing. To simply deny that certain capacities exist, despite their presence in our everyday lives, is nearly the opposite of scientific practice. Indeed those capacities we are calling normative underlay scientific practice itself.

We begin with the fact that our social practices are thoroughly normative—normative “all the way down,” so to speak. Inferential moves, decisions made in neuroscience laboratories, attributions of guilt or innocence in courts of law, our expectations about the conduct of our friends and neighbors, our expectations of our own conduct: these are all normative practices, and they occur in what we will call “normative contexts.” We hold people responsible for their actions in such normative contexts: when they follow the rules, they are praiseworthy, or at least not blameworthy. But to say a person is praiseworthy or blameworthy is to say, not simply that she acted in accordance with the rules, but that she exercised *control* over her conduct. This feature of our experience is linked to our capacity to reason about our *practical* engagements. Responsibility-attributions are fundamental to those practices, and to give them up, even if it were possible, would almost certainly dramatically alter our social experience.

Determinism, roughly, claims the following:

1. Every event has an antecedent cause.
2. All human acts are events.
3. Therefore, all human acts are caused by antecedent events.
4. If events are caused, they are causally determined, or in some sense *necessary*.
5. Therefore, human beings are not free to choose an act from an array of possible acts.

Contemporary scientific determinism holds that the antecedents of acts include neural events over which the agents of the acts have no control.

However, causal relationships are ubiquitous, not only in science but in our daily lives, including in our attributions of responsibility. Even if much of our behavior is caused by neural events that are modeled with no reference to the sort of control that seems to ground responsibility, much of our behavior can only be described in terms of things we do. As Cartwright (2007) argues, nature is rich with types of causal relation, and both natural and specialized scientific languages include ineliminable causal verbs. Among the causal accounts of human behavior are narratives about the formation of intentions and the

embedding of those intentions in plans. Why did Bill pull the trigger of the gun that resulted in Judy's death? At the most general level, Jones may have intended to do so, and may even have formulated a detailed plan for bringing about Judy's death. There may, however, be evidence, perhaps including fMRI, of neural functioning that falls outside species typical functioning (i.e. is deviant), such as significantly reduced serotonin levels in the brain. If the former is true, we can hold Bill fully *accountable* for his behavior. If the latter is true, justice may require that we remove Bill from society until he is *capable of controlling* his aggression. Between these two possibilities, Bill's neural deviance may only be sufficient to mitigate, but not fully *excuse*, his conduct. These are all causal stories and fit the claims of determinism 1–5 above. Indeed, there may also be neural events that must be included in the causal stories we tell about degrees of control, having intentions, and making plans, if we wish to flesh out our explanations of human conduct.

Current versions of compatibilism generally emphasize the human capacities that ground attributions of responsibility, without postulating the existence of entities or systems that fail to behave in accordance with causal laws, including laws relating neural events to behavior. Current versions of compatibilism emphasize the hierarchical structure of decision-making (Frankfurt, 2006); so-called “reactive attitudes,” such as resentment that would make no sense in the absence of responsibility attributions (Strawson, 1962); the capacity to make plans that provides the practical context of intentional acts (Bratman, 2007); the practical necessity of adopting a responsibility-attributing stance toward certain types of system whose behavior we need to predict (Dennett, 1987); or the weakest form of control that can ground responsibility attributions (Fischer & Ravizza, 1998). Common to all of these views is the claim that, to the greatest extent possible, theories of responsibility must preserve as many of the characteristics of personhood as possible, compatible with the best scientific accounts of human behavior.

We cannot here examine in detail the bewilderingly complex literature on compatibilism. Our position is broadly compatibilist, and draws primarily on the work by Fischer & Ravizza (1998), which emphasizes the role of control in grounding responsibility-attributions. We also, like Dennett, believe that the starting point in addressing the problem posed by viewing human conduct as causally determined should be the pragmatic one of determining the weakest set of requirements of our normative practices. We shall, however, suggest that those practices are not fixed and static. They change with new scientific discoveries about our capacities. One such discovery is that there are neural processes that underlie our capacity to initiate events, that is, to be the causal *source* of events, and not simply an object in a billiard-ball model of causality. Agency is grounded on bodily events, and especially neural events, rather than being an alternative to the causal structure of the world.

7. Drawing distinctions: clearing space for moral responsibility

Folk-psychological notions of free will and moral responsibility continue to hold sway not only over moral attributions of praise- and blame-worthiness, including the legal notion of culpability, but also over the philosophical inferences drawn from neurobiological discoveries by scientists like Green and Cohen. Broadly, the idea is that holding people responsible for their conduct entails that when a person acts, she is at least in principle able to do otherwise (Frankfurt, 1988). If our actions are determined by events outside our control, we cannot be held responsible. Traditionally, this has been interpreted to mean that there is a conflict between free will and determinism. A few distinctions, common in philosophy but not in neuroscience, will show that the capacities that underlay our normative, including our moral, practices, do not presuppose a metaphysical notion of free will.

As Fischer and Ravizza (1998) have shown, all that is required for attributions of responsibility is the capacity to control one's behavior. Control is not a mechanism that issues from a "free will," and it is perfectly compatible with determinism. Physical systems can include a mechanism for controlling the system, within certain constraints. A home heating system includes a thermostat that modulates temperature. We do not, of course, hold heating systems "responsible" for changes in room temperature. Fischer and Ravizza, however, suggest a more complex conception of control that explains at least part of what we mean when we assert that a person is responsible for her actions. We will argue that Fischer and Ravizza do not solve the problem posed by determinism, but they lay the foundation for a tolerable solution.

Fischer and Ravizza distinguish between two sorts of control: regulatory control and guidance control. Regulatory control is the capacity to choose from any one of an array of actions. Guidance control is the capacity to perform an action in a certain manner, including the capacity to refrain from performing the action. Even where regulatory control is impossible, because of environmental or physiological constraints, guidance control may be operative. Thus, a person with a physiological limitation, such as blindness, may have a very limited array of possible acts to choose from, but within the range of possible acts she can usually select a manner of performing them (at least with practice). Her actions are determined in the sense that she is not free to choose to perform every act sighted people can perform. But she can guide her behavior to achieve certain ends, she can plan a course of action, and she can choose not to take that course of action.

Fischer and Ravizza argue that guidance control is all that is required for a robust notion of responsibility that can ground our norm-ascribing practices. The conception of responsibility they have in mind is what they call "reason-responsiveness." We can give reasons for our actions and understand other people's reasons for their actions, within the constraints set by the environment, including physical laws, and our physiological capacities. At least some of the causes of our actions are normative in the broad sense: we can follow rules; we can make mistakes; we can act on our desires. The blind person can choose to harm or not harm a person, within the physiological constraints set by her inability to see. For those choices, that person can be held responsible. She can understand and follow a substantial range of proprieties of practice.

There is a problem with this analysis, to which we return below: what is it that is responsive to reasons? After all, thermostats are responsive to the difference between the optimal temperature of a room and the actual temperature, yet we are not tempted to adopt an "intentional stance" toward heating systems and hold them responsible for temperature modulation (Dennett, 1987). But first, let us consider the extent to which the account of responsibility of Fischer and Ravizza circumvents the claim that the causal role played by serotonin levels is determinist in a way that undermines a conception of responsibility that can ground our ascriptions of responsibility and culpability. To be sure, serotonin levels, other things being equal, place constraints on the range of choices over which a person can select a course of action. A depletion of serotonin is not the only constraint, and other neurochemical systems may play a role in modulating the impact of serotonin levels on behavior, but there is nothing intentional about such constraints. They seem not to be susceptible to the influence of reasons. But if Fischer and Ravizza are correct, then even if the range of choices is constrained by serotonin levels and reuptake efficiency, so that depleted serotonin or interference with its binding to neural receptors is likely to produce aggressive behavior, the expression of aggression may nonetheless be within the control of the agent. The capacity to act aggressively certainly conferred an advantage on early hominids, and random mutations in serotonin production mechanisms that promoted aggressive behavior survived for that reason. But if aggressive impulses were not subject to

guidance control, the result would probably have been disastrous for the survival of *homo sapiens sapiens*. So guidance control mechanisms also conferred a reproductive advantage.

To clarify the way reasons play a role in the exercise of guidance control, consider the case of a neuroscience professor with a history of flying into a rage whenever her aims are thwarted. Most of the time she either expresses her rage verbally or suppresses her rage when it seems prudent to do so. However, on one occasion, after she has been denied tenure, she borrows a gun, she practices at a firing range for a few weeks, and she brings the gun to a biology department meeting. She sits quietly through most of the meeting, and then she pulls the gun from her purse and starts to shoot her colleagues. It is not too far-fetched, we believe, to conjecture that certain neural processes constrain the range of alternatives available to this woman for expressing her rage. She has limited regulatory control over her conduct. However, within that range of alternatives, those biological processes are themselves reason-responsive. Indeed, the professor has formulated her intention to shoot the people she holds responsible for her academic failure, she formulates a plan within which that intention is embedded, and she carries out the plan in a perfectly reasonable way, given her aims. She may not be able, because of neurochemical processes involving serotonin levels (or some other model of aggression), to refrain from expressing her rage, but she apparently can express her rage in a particular manner for specific reasons. Even if her attorney at trial presents an expert neuroscientist to explain how her low serotonin levels played a significant causal role in her violent conduct, it is quite likely the jury will recognize that she had sufficient control to hold her responsible, although the same expert opinion may be the basis for applying a mitigating factor at sentencing. There is no reason to believe neurochemical processes cannot be reason-responsive.

The problem of responsibility in a deterministic world remains, however. It is simply pushed up one level. Surely reason-responsiveness is itself executed by neurochemical events, although we do not yet have detailed models of the neural basis of such normative capacities. We do, however, have a place to start developing such models: the prefrontal cortex.

8. An alternative view: towards a theory of the neural basis of reason-responsiveness with respect to the control of impulsive aggressive rage behavior

We have argued that reason-responsiveness can ground responsibility even if impulsive aggressive behavior has a neural basis. But reason-responsiveness itself is a human capacity that, as determinists, we believe should have a neurochemical basis. The issue here is whether there are neurochemical models that can explain the reason-responsiveness of guidance control. We sketch a possible model in the next section.

8.1. Essential characteristics of the neural substrate

The key to this model is that the principal structure(s) involved must serve as a substrate for classical conditioning that will underlie deciding on courses of action on the basis of reasons. At a minimum, we must identify a structure, a set of structures, or region(s) that has the capacity: (1) to receive inputs from sensory and association areas of the brain; (2) for establishing classical conditioning of neurons; and (3) to regulate and inhibit the regions of the brain that integrate the expression of the affective (impulsive) rage response. Moreover, the model must (4) account for the experience of acting *for reasons*. If we cannot control our aggressive impulses by engaging in certain normative procedures, including reasoning, we can hardly be held responsible for those of our aggressive acts that cause harm. If there is

anything left of the concept of “free will,” it is this: most of the time we act *for reasons*, even when we perform bad acts for bad reasons.

One likely candidate that contains all of these properties is the prefrontal cortex. Its uniqueness is underscored, in part, by its anatomical relationships with other regions of the cortex as well as with the near and distant structures which its axons target. Specifically, the prefrontal cortex acts as the ‘hub’ of the cerebral cortex in that it receives inputs from all regions of the cortex, including all regions associated with reception of all types of sensory signals initiated from the periphery (Rolls, 2004). In addition, it receives important inputs from key nuclei of the limbic system such as the amygdala and hippocampal formation, as well as the medio-dorsal nucleus of the thalamus (Cavada and Reinos-Suarez, 1985; Fuster, 2001; Price, 2005). Thus, the sensory and other inputs to the prefrontal cortex provide one of the critical features essential for a conditioning process to take place within the prefrontal cortex.

A second feature is the conditioning process of neurons in the prefrontal cortex. In classical conditioning, a neutral stimulus is paired with either an aversive or positively reinforcing stimulus. After such pairings are repeated a sufficient number of times, the previously neutral stimulus takes on the properties of the negatively or positively reinforcing stimulus and the organism (animal or human) responds to the previously neutral stimulus with the same type of emotion as it did to the aversive or positively reinforcing stimulus. The neural basis for such conditioning of neurons in the central nervous system in general and in the prefrontal cortex in particular has been demonstrated (Rolls, 2005). In brief, utilizing a conditioning paradigm, it has been shown that synaptic efficacy is enhanced at synapses where stimulus inputs that were previously neutral have now taken on emotional properties for the organism. Enhancement of synaptic connections under these conditions may also reflect the development of new synapses as well. As noted above, conditioning and new synapse development within the prefrontal cortex is facilitated by the fact that it receives inputs from such a wide variety of sensory and other regions such as the association cortex. Accordingly, neural events underlying the conditioning process within the prefrontal cortex exhibits the second feature described above.

The third feature requires an output pathway from the region in question to the hypothalamus and PAG from which impulsive rage behavior is integrated and expressed. Here, too, the prefrontal cortex satisfies this criterion. Prefrontal cortical regulation of aggressive behavior has now been well established (Siegel, Edinger and Lowenthal, 1974; Raine, Buchsbaum, and LaCasse, 1997; Raine, Lencz, Bihrie, LaCasse, Colletti, 2000; Raine, Meloy, Bihrie, Stoddard, La Casse, Buchsbaum, Pietrini, Guazzelli, Basso, Jaffe, and Grafman, 1998). Prefrontal cortical control of aggression may be manifest through fibers originating from the prefrontal cortex that can reach the hypothalamus directly (Öngur, An, and Price, 1998), or indirectly, through an initial projection to the medio-dorsal thalamic nucleus and from this nucleus to the hypothalamus via a series of interneurons in the midline thalamus (Siegel, Edinger, and Troiano, 1973). Prefrontal cortical control of aggression can also occur via direct anatomical projections from the prefrontal cortex to the PAG (An, Bandler, Öngur, and Price, 1998). Thus, we can see that the prefrontal cortex fulfills three of the conditions we set above of a structure or region which could serve as the substrate for making reason-responsive decisions.

8.2. A description of how this model may actually function and its implications for establishing responsibility for one’s actions

Can the system herein described satisfy the fourth condition: is it reason-responsive? Consider first the response of a highly impulsive individual who is irritated by the actions of another person who cuts in front of her while driving on the highway or who insults her at a

public meeting. Typically, the response of such an individual would likely be to angrily confront the person perceived as intolerably rude or insulting. The position advocated here is that there would be little involvement of the impulsive individual's prefrontal cortex. In addition, it is possible that such an individual with impulsive tendencies might have lower than normal levels of circulating brain serotonin (as described earlier in this paper). Thus, in the absence of inhibitory mechanisms from the prefrontal cortex to inhibit aggression integrated within the hypothalamus and PAG and with possible reduced activation of serotonin receptors in these regions, the impulsive, aggressive response system will likely outweigh any forces present to suppress these aggressive tendencies.

Now consider another scenario, one involving an individual in which the types of confrontation are the same as in the case above, but where the history of the individual targeted by the confrontations differ. Suppose that the individual had been exposed to a series of lectures and reading material concerning anger management. Here, the education involved a conditioning-learning process in which the individual is sensitized to sudden increases in bodily and emotional tension resulting from a confrontational situation. The individual is further taught to associate such autonomic and emotional reactions with the futility of reacting to the confrontation with anger or aggression. After a period of training, the individual is then exposed to a real-life confrontational event. What might this individual do? If the training has been effective, the individual may initially begin to show anger in the presence of the confrontational stimulus. But within a fraction of a second, the initiation of anger triggers an association with the thought of the futility of expressing an angry or aggressive response, resulting in suppression of an aggressive response.

In the second scenario, it seems natural to suppose that the educational initiative enables the person to consider reasons for refraining from acting out violently. The associationist psychological terminology is somewhat misleading, but the causal role of reasons in the suppression of aggression is apparent: the educational initiative may have affected the person's neurochemistry, but she is capable of recognizing that a violent response would be *futile*, and she can guide her behavior accordingly.

Under a folk-psychological description, the individual is responding to a variety of reasons for restraining her aggressive behavior. While she may not have regulatory control, because her neural response system triggers anger, she has guidance control as a result of a kind of moral education. Under a neurobiological description, sensory signals associated with a confrontational event pass from auditory and visual cortices to the prefrontal cortex. As a result of a process involving the development and strengthening of synaptic connections within the prefrontal cortex, a neural conditioning process has taken place within the prefrontal cortex. This neural conditioning process underlies its behavioral counterpart of the thought process linking the increase in anger and emotional tension with the counter notion of the futility that would result from the expression of anger or aggression in response to the confrontational situation. Finally, activation of the prefrontal cortex results in excitation of the 'inhibitory' output pathways from the prefrontal cortex to the hypothalamus and PAG, thus inhibiting the rage mechanism. The circuitry depicting this control mechanism is shown in Fig. 1. We see no conflict between describing the person's choice not to respond aggressively as reason-responsive, and therefore under her guidance control, and describing the response as caused by neural inhibiting processes that occur in the prefrontal cortex.

Now, there are almost certainly degrees of prefrontal cortical involvement in human conduct, including the expression of aggression. In the first scenario, the assumption was that little cortical control occurred, but in fact exactly the same aggressive behavior might have been the result of the "blink" phenomenon or rapid cognition (Gladwell, 2005). Or it

might have been the result of more careful reasoning. In other words, one can act badly *for reasons*, as presumably was the case in the biology professor example. As Frankfurt (2006, p. 21) points out, “The normative authority of reason ... cannot be what accounts for the normative authority of morality. ... People who behave immorally incur a distinctive kind of opprobrium. ... Our response to sinners is not the same as our response to fools.” Acting badly can be reason-responsive, and therefore subject to blame or punishment. Linking punishment to strategies for improving one’s reasoning abilities, or at least for making them less antisocial, may be the best response of the criminal justice system to the new neuroscience.

9. Becoming responsible

As any parent is likely to agree, acting responsibly is the result of moral education: learning to be reason-responsive, and learning to respond specifically to reasons for restraint. We may overvalue the role of consciousness in voluntary action, as George Sher has argued recently (Sher, 2009), but becoming conscious of reasons for *not* acting inappropriately is nonetheless a key to the minimal conditions for holding people responsible for their actions. In this context, it is worth considering Benjamin Libet’s solution to the problem of free will in light of his famous experiments on the delay from an initial neural process that terminates in an act and the conscious willing of the act. Briefly, if tests subjects are told to perform an act, e.g. to flex their wrists, often and at times of their own choosing, is the act caused by an antecedent neural process, a conscious decision, or both? By attaching electrodes to the a subject’s brain and wrists, and telling him to notice as precisely as possible when he consciously decided to flex, it appeared that the neural process (the “readiness potential” or RP) begins about 350 milliseconds before the decision to act, and the decision to act occurs about 200 milliseconds before the act. The implications of this experiment were obvious: if an unconscious neural process that terminates in an act begins prior to the decision to act, then how is it that the act is the result of the decision?

Libet’s own answer to this philosophical question resembles our suggestion that even persons with a genetically coded or environmentally produced depletion of brain serotonin levels may have the capacity to refrain from acting aggressively. Libet argued that even if consciousness did not, or could not, initiate an act, an agent might be able to exercise a “conscious veto” of the act. Indeed, when a subject attempted to abort the wrist flexion act, his RP began normally, but then flattened out about 200 milliseconds before the act was scheduled to occur. Libet’s experiments may seem crude by today’s standards, but they have been replicated by other scientists, and still represent a major empirical argument against free will, if free will is understood to be the capacity to make conscious decisions about which actions to perform. Indeed, fMRI has been used to demonstrate that decisions can be predicted up to ten seconds before subjects become conscious of their decisions (Chun Siong Soon, Marcel Brass, Hans-Jochen Heinze and John-Dylan Haynes, 2008). If, however, free will is simply the capacity to exercise a conscious veto, Libet’s experiments provide empirical support for a version of compatibilism.

If Libet’s solution is correct, then the sort of moral education we discuss above should be at the forefront of our efforts to reduce the amount of unnecessary violence in the world, or at least in the corner of the world in which we can provide that sort of education. We *become* responsible not for initiating actions, but for exercising a veto over proscribed actions.

10. Conclusions

The person described in the second scenario above has a clear choice between expressing anger and aggression or responding to reasons for exercising control over her conduct. She

may, of course, decide to carry out an aggressive response, but that decision, too, is reason-responsive. The person has a choice whether to respond violently, a choice which is governed by clear neural mechanisms and which could be regulated by behavioral conditioning, appropriate religious and other ethical education, and training. This gets us back to the initial question: is a person responsible for her actions or do pre-existing neural circuits and neurochemical levels of key neurotransmitters within her brain determine her actions?

In an attempt to answer to this question, it is reasonable to conclude that, in a wide variety of circumstances, it seems that a person has the capacity to control her behavior, within the constraints set by biology and environment. She can develop conditioning or related training procedures, perhaps out of conscious awareness of her propensity to express anger or aggression, that endow her with the ability to exercise a veto over impulsive aggressive actions. From this view, the claim that "her brain neurochemistry made her do it" ignores what we already know about the brain's capacities. That is an ironic implication of the claim. But, what if a person's brain neurochemistry does drive her to act violently? Is this statement incompatible with what was said above? Not necessarily. There may well be individuals who, for example, have very low levels of brain serotonin (and possibly abnormal levels of other neurotransmitters that regulate aggressive behavior) which, for unknown reasons, are resistant to pharmacological treatment, psychotherapy, and other behavioral conditioning procedures. Such individuals may not have the degree of guidance control to warrant attribution of responsibility and culpability. They may more closely resemble the Parkinson patient who suffers from a hypokinetic disorder. Therefore, in evaluating the behavior of an individual in terms of assessing whether she is responsible for his actions, the extent to which an individual is capable of conscious deciding not to perform an inappropriate act must be factored into the category in which she is placed, and the appropriate response of the criminal justice system.

Acknowledgments

Portions of this research were supported by NIH NS 07941-36. Portions of this paper were presented at the biennial meeting of the 31st Congress of the International Academy of Law and Mental Health in New York City, June, 2009.

References

- An X, Bandler R, Öngür D, Price JL. Prefrontal cortical projections to longitudinal columns in the midbrain periaqueductal gray in macaque monkeys. *Journal of Comparative Neurology* 1998;401:455–479. [PubMed: 9826273]
- Asberg M, Traskman L, Thoren P. 5-HIAA in the cerebrospinal fluid. A biochemical suicide predictor? *Archives of General Psychiatry* 1976;33:1193–1197. [PubMed: 971028]
- Badawy AA. Alcohol, aggression and serotonin: metabolic effects. *Alcohol and Alcoholism* 1998;33:66–72. [PubMed: 9539179]
- Bhatt S, Bhatt RS, Zalcman SS, Siegel A. Peripheral and central mediators of lipopolysaccharide induced suppression of defensive rage behavior in the cat. *Neuroscience* 2009;163:1002–1011. [PubMed: 19647047]
- Bratman, ME. *Structures of Agency*. New York: Oxford; 2007.
- Brunton, LL.; Lazo, JS.; Parker, KL. *Goodman and Gilman's "The pharmacological basis of therapeutics"*. New York: McGraw Hill; 2006.
- Cartwright, N. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press; 1999. Ch. 5.
- Cartwright, N. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge: Cambridge University Press; 2007. p. 19-20, 43-53.

- Cavada C, Reinoso-Suarez F. Topographical organization of the cortical afferent connections of the prefrontal cortex in the cat. *Journal of Comparative Neurology* 1985;242:293–324. [PubMed: 2418073]
- Coccaro EF, Kavoussi RJ, Trestman RL, Gabriel SM, Cooper TB, Siever LJ. Serotonin function in human subjects: intercorrelations among central 5-HT indices and aggressiveness. *Psychiatry Research* 1997;73:1–14. [PubMed: 9463834]
- Coccaro EF, Kavoussi RJ. Fluoxetine and impulsive aggressive behavior in personality-disordered subjects. *Archives of General Psychiatry* 1997;54:1081–1088. [PubMed: 9400343]
- Coccaro EF, Kavoussi RJ, Cooper TB, Hauger RL. Central serotonin activity and aggression: Inverse relationship with prolactin response to *d*-fenfluramine, but not CSF 5-HIAA concentration, in human subjects. *American Journal of Psychiatry* 1997;154:1430–1435. [PubMed: 9326827]
- Constantino JN, Morris JA, Murphy DL. CSF 5-HIAA and Family History of Antisocial Personality Disorder in Newborns. *American Journal of Psychiatry* 1997;154:1771–1773. [PubMed: 9396964]
- Cremniter D, Jamain S, Kollenbach K, Alvarez JC, Lecrubier Y, Gilton A, Jullien P, Lesieur P, Bonnet F, Spreux-Varoquaux O. CSF 5-HIAA levels are lower in impulsive as compared to nonimpulsive violent suicide attempters and control subjects. *Biological Psychiatry* 1999;45:1572–1579. [PubMed: 10376117]
- Daderman AM, Lidberg L. Relapse in violent crime in relation to cerebrospinal fluid monoamine metabolites (5-HIAA, HVA and HMPG) in male forensic psychiatric patients convicted of murder: a 16-year follow-up. *Acta Psychiatrica Scandinavica Supplement* 2002;412:71–74.
- Dennett, D. *The Intentional Stance*. Cambridge, M.A.: M.I.T. Press; 1987.
- Dolan MC, Anderson IM. The relationship between serotonergic function and the Psychopathy Checklist: Screening Version. *Journal of Psychopharmacology* 2003;17:216–222. [PubMed: 12870570]
- Dupre J. The solution to the problem of freedom of the will. *Noûs* 1996;Vol. 30:385–402. Supplement: Philosophical Perspectives 10, Metaphysics, 1996 (1996).
- Echizen H, Freed CR. Measurement of serotonin turnover rate in rat dorsal raphe nucleus by *in vivo* electrochemistry. *Journal of Neurochemistry* 2006;42:1483–1486. [PubMed: 6200572]
- Fairbanks LA, Melega WP, Jorgensen MJ, Kaplan JR, McGuire MT. Social impulsivity inversely associated with CSF 5-HIAA and fluoxetine exposure in vervet monkeys. *Neuropsychopharmacology* 2001;24:370–378. [PubMed: 11182532]
- Fischer, JM.; Ravizza, MSJ. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge U.K.: The Cambridge University Press; 1998.
- Frankfurt, H. The Tanner Lectures in Moral Philosophy. In: Debra, Satz, editor. *Taking Ourselves Seriously*. Stanford: Stanford University Press; 2006. p. 1-26.
- Frokjaer VG, Mortensen EL, Nielsen FA, Haugbol S, Pinborg LH, Adams KH, Svarer C, Hasselbalch SG, Holm S, Paulson OB, Knudsen GM. Frontolimbic serotonin 2A receptor binding in healthy subjects is associated with personality risk factors for affective disorder. *Biological Psychiatry* 2008;63:569–576. [PubMed: 17884017]
- Fuchs SAG, Edinger HM, Siegel A. The organization of the hypothalamic pathways mediating affective defense behavior in the cat. *Brain Research* 1985;330:77–92. [PubMed: 4039213]
- Fuster J. The Prefrontal Cortex--An Update: Time Is of the Essence. *Neuron* 2001;30:319–333. [PubMed: 11394996]
- Green J, Cohen J. For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London. B* 2004;1775–1785.
- Hassanain M, Bhatt S, Siegel A. Differential modulation of feline defensive rage behavior in the medial hypothalamus by 5-HT1A and 5-HT2 receptors. *Brain Research* 2003;981:201–209. [PubMed: 12885442]
- Hibbeln JR, Umhau JC, Linnoila M, George DT, Ragan PW, Shoaf SE, Vaughan MR, Rawlings R, Salem N Jr. A replication study of violent and nonviolent subjects: cerebrospinal fluid metabolites of serotonin and dopamine are predicted by plasma essential fatty acids. *Biological Psychiatry* 1998;44:243–249. [PubMed: 9715355]

- Kim-Cohen J, Caspi A, Taylor A, Williams B, Newcombe R, Craig IW, Moffitt TE. MAOA, maltreatment, and gene-environment interaction predicting children's mental health: new evidence and a meta-analysis. *Molecular Psychiatry* 2006;11:903–913. [PubMed: 16801953]
- Krakowski M. Violence and Serotonin: Influence of Impulse Control, Affect Regulation, and Social Functioning. *Journal of Neuropsychiatry and Clinical Neuroscience* 2003;15:294–305.
- Libet B. Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioral and Brain Sciences* 1985;8:529–539.
- Lidberg L, Belfrage H, Bertilsson L, Evenden MM, Asberg M. Suicide attempts and impulse control disorder are related to low cerebrospinal fluid 5-HIAA in mentally disordered violent offenders. *Acta Psychiatrica Scandinavica* 2000;101:395–402. [PubMed: 10823300]
- Lidberg L, Daderman A. Reduced serotonin levels are predisposing to violence. A simple blood test predicts dangerous character. *Lakartidningen* 1997;94:3385–3388. [PubMed: 9379806]
- Mehlman PT, Higley JD, Faucher I, Lilly AA, Taub DM, Vickers J, Suomi SJ, Linnoila M. Low CSF 5-HIAA concentrations and severe aggression and impaired impulse control in nonhuman primates. *American Journal of Psychiatry* 1994;151:1485–1491. [PubMed: 7522411]
- Moeller FG, Dougherty DM, Swann AC, Collins D, Davis CM, Cherek DR. Tryptophan depletion and aggressive responding in healthy males. *Psychopharmacology (Berl)* 1996;126:97–103. [PubMed: 8856827]
- Öngür D, An X, Price JL. Prefrontal cortical projections to the hypothalamus in macaque monkeys. *Journal of Comparative Neurology* 1998;401:480–505. [PubMed: 9826274]
- Pihl RO, Young SN, Harden P, Plotnick S, Chamberlain B, Ervin FR. Acute effect of altered tryptophan levels and alcohol on aggression in normal human males. *Psychopharmacology (Berl)* 1995;119:353–360. [PubMed: 7480513]
- Pihl RO, Le Marquand D. Serotonin and aggression and the alcohol-aggression relationship. *Alcohol and Alcoholism* 1998;33:55–65. [PubMed: 9539178]
- Price JL. Free will versus survival: Brain systems that underlie intrinsic constraints on behavior. *Journal of Comparative Neurology* 2005;493:132–139. [PubMed: 16255003]
- Raab A. A rapid and sensitive method for the determination of serotonin turnover in discrete brain areas. (1970). *Journal of Comparative Physiology* 1970;68:272–275.
- Raine A, Buchsbaum M, LaCasse L. Brain abnormalities in murderers indicated by positron emission tomography. *Biological Psychiatry* 1997;42:495–508. [PubMed: 9285085]
- Raine A, Meloy JR, Bihrie J, Stoddard L, La Casse L, Buchsbaum M. Reduced prefrontal and increased subcortical brain functioning assessed using positron emission tomography in affective and predatory murderers. *Behavioral Sciences and the Law* 1998;16:329–332.
- Raine A, Lencz T, Bihrie S, LaCasse L, Colletti P. Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder. *Archives of General Psychiatry* 2000;57:119–127. [PubMed: 10665614]
- Reif A, Rosler M, Freitag CM, Schneider M, Eujen A, Kissling C, Wenzler D, Jacob CP, RetzJunginger P, Thome J, Lesch KP, Retz W. Nature and nurture predispose to violent behavior: Serotonergic genes and adverse childhood environment. *Neuropsychopharmacology* 2007;32:2375–2383. [PubMed: 17342170]
- Rolls ET. The functions of the orbitofrontal cortex. *Brain and Cognition* 2004;55:11–29. [PubMed: 15134840]
- Saudou F, Amara DA, Dierich A, Lemeur M, Ramboz S, Segu L, Buhot MC, Hen R. Enhanced Aggressive behavior in mice lacking the 5HT_{1B} receptor. *Science* 1994;265:1875–1878. [PubMed: 8091214]
- Shaikh MB, Barrett JA, Siegel A. The pathways mediating affective defense and quiet biting attack behavior from the midbrain central gray of the cat: an autoradiographic study. *Brain Research* 1987;437:9–25. [PubMed: 3427485]
- Shaikh MB, De Lanerolle NC, Siegel A. Serotonin 5-HT_{1A} and 5-HT_{2/1C} receptors in the midbrain periaqueductal gray differentially modulate defensive rage behavior elicited from the medial hypothalamus of the cat. *Brain Research* 1997;765:198–207. [PubMed: 9313892]
- Sher, G. *Who Knew: Responsibility Without Awareness*. Oxford: Oxford University Press; 2009.

- Siegel A, Bhatt S, Bhatt R, Zalcman SS. The neurobiological bases for development of pharmacological treatments of aggressive disorders. *Current Neuropharmacology* 2007;5:135–147.
- Siegel A, Edinger H, Troiano R. The pathway from the mediodorsal nucleus of thalamus to the hypothalamus in the cat. *Experimental Neurology* 1973;38:202–217. [PubMed: 4347813]
- Siegel A, Edinger H, Dotto M. Effects of electrical stimulation of the lateral aspect of the prefrontal cortex upon attack behavior in cats. *Brain Research* 1975;93:473–484. [PubMed: 1236761]
- Siegel A, Roeling TAP, Gregg TR, Kruk MR. Neuropharmacology of brain-stimulation-evoked aggression. *Neuroscience and Biobehavioral Reviews* 1999;23:359–389. [PubMed: 9989425]
- Siegel, A. *The neurobiology of aggression and rage*. Boca Raton, FL: CRC Press; 2005.
- Siegel A, Victoroff J. Understanding human aggression: New insights from neuroscience. *International Journal of Law and Psychiatry* 2007;32:209–215. [PubMed: 19596153]
- Soon CS, Brass M, Heinze H-J, Haynes J-D. Unconscious determinants of free decisions. *Nature Neuroscience* 2008;11:543–545.
- Stanley B, Molcho A, Stanley M, Winchel R, Gameroff MJ, Parsons B, Mann JJ. Association of aggressive behavior with altered serotonergic function in patients who are not suicidal. *American Journal of Psychiatry* 2000;157:609–614. [PubMed: 10739421]
- Strawson P. Freedom and Resentment. *Proceedings of the British Academy* 1962;48:187–211.
- Williams WA, Shoaf SE, Hommer, Linnoila M. Effects of acute tryptophan depletion on plasma and cerebrospinal fluid tryptophan and 5-hydroxyindoleacetic acid in normal volunteers. *Journal of Neurochemistry* 1999;72:1641–1647. [PubMed: 10098872]

Fig. 1.

Schematic diagram illustrating the role of the prefrontal cortex (PFC) as a key site in the conditioning process by which this region of cerebral cortex regulates and controls aggressive impulsive behavior that is mediated by the medial hypothalamus and midbrain periaqueductal gray (PAG). The PFC receives major inputs from sensory regions of cortex and thalamus. These inputs provide the basis by which classical conditioning of neurons in this region takes place and accordingly provide the substrate for cognitive associations to be formed in this region. Activation of these cognitive associations is associated with activation of PFC neurons which then suppress medial hypothalamic and PAG mechanisms, thus reducing or eliminating the expression of the impulsive rage response.