International Journal of Law and Psychiatry 32 (2009) 189–197

Contents lists available at ScienceDirect

International Journal of Law and Psychiatry

Introduction

# Aggression, science, and law: The origins framework

Jeff Victoroff *

*University of Southern California Keck School of Medicine, Departments of Neurology and Psychiatry, 7601 E. Imperial Highway, Downey, CA 90242, USA*

ARTICLE INFO

*Keywords:*
Aggression
Evolution
Neuroscience
Law
Free will
Criminal responsibility

ABSTRACT

Human societies have formalized instincts for compliance with reciprocal altruism in laws that sanction some aggression and not other aggression. Neuroscience makes steady advances toward measurements of various aspects of brain function pertinent to the aggressive behaviors that laws are designed to regulate. Consciousness, free will, rationality, intent, reality testing, empathy, moral reasoning, and capacity for self-control are somewhat subject to empirical assessment. The question becomes: how should law accommodate the wealth of information regarding these elements of mind that the science of aggression increasingly makes available? This essay discusses the evolutionary purpose of aggression, the evolutionary purpose of law, the problematic assumptions of the *mens rea* doctrine, and the prospects for applying the neuroscience of aggression toward the goal of equal justice for unequal minds. Nine other essays are introduced, demonstrating how each of them fits into the framework of the permanent debate about neuroscience and justice. It is concluded that advances in the science of human aggression will have vital, but biologically limited, impact on the provision of justice.



Aggression is normal, natural behavior throughout the kingdom of animals. Aggression may be defined as behavior that serves to control another organism. This definition may seem broad, excluding as it does the classic components of intent and harm. But ethologists, even microbiologists, routinely observe creatures without brains or spines or even bilaterally symmetrical body plans approaching and displacing or eating others. "Intent"—a popular (and often misunderstood) concept in Western law—is hardly at issue in brainless creatures. But no observer could dispute the appearance of aggressive behavior, and no biologically savvy observer would dispute the notion that competitive or defensive or predatory behaviors represent teleologically identical products of evolutionary selection, whether that behavior is witnessed in sea snails, prairie dogs, or presidents. And "harm" is simply incorrect, with or without the added conflation of intentionality, when one considers aggressive behaviors such as the alpha animal's rough actions to maintain a mutually beneficial hierarchy or the loving parent's shout at and vigorous pulling apart of fighting toddlers. Any definition of aggression will be hotly debated. I submit, however, that a broad definition rooted in evolutionary meaning ultimately illuminates otherwise hard-to-account for actions of social animals.

Aggression has been classified, in rodents, as conforming to one or another of a few categories of evolutionary purpose: predation, defense, territorial, dominance related, maternal, and sexual (e.g., Moyer, 1976). It is virtually certain that the genes and pieces of nervous systems that evolved to mediate these behaviors in animals with smaller brains were modified and incorporated into the genes and brain parts of humans (e.g., Butler & Hodos, 2005; Eccles, 1989; MacLean, 1990; Mithen, 1996; Reep, Finlay, & Darlington, 2007; Scheibel & Schopf, 1997). But one must be wary of one-to-one matching. Aggression has also been classified according to a dichotomy that claims polar differences between offensive/predatory/premeditated/instrumental versus defensive/reactive/hostile/affective acts (e.g., Feschbach, 1964; Geen, 2001; Kingsbury, Lambert, & Hendrickse, 1997; McEllistrem, 2004). Yet this dichotomy, rooted in observations of nonhuman mammals, is conceptually weak (e.g., Bushman & Anderson, 2001; Little, Brauner, Jones, Nock, & Hawley, 2003; Parrotta & Giancola, 2007). It strives to dichotomize human behaviors that are multidetermined, simultaneously serve affective and instrumental purposes, and probably recruit cerebral regions that may have been phylogenetically dedicated to one or another pure type of aggression, but that have been adopted into extremely complex cortical–subcortical circuits. The soldier on patrol in a free-fire zone is probably using brain parts designed for predation, including stalking, pouncing, and killing. So is the serial rapist looking for targets. Neither man will eat what he attacks. So, even if the cerebral circuitry and neurochemistry invoked greatly overlaps with that which supports the success of the lioness, the essential purpose of the men's complex behavior is very different from the calorie/protein-seeking of any other predator. And the hunting soldier who fires at a threat is probably employing circuitry evolved to mediate attack, defense, approach, avoidance, and affect. The General Aggression Model of Anderson and Bushman (2002) comes closer to describing real-life events, recognizing as it does that human acts are often *both* instrumental and affect-based.

* Tel.: +1 310 316 0761; fax: +1 310 540 6785.
*E-mail address:* victorof@usc.edu.

0160-2527/$ – see front matter © 2009 Published by Elsevier Ltd.
doi:10.1016/j.ijlp.2009.04.009

Science has made considerable progress in identifying the neurobiological correlates of aggression. In animals with brains combining cortical and subcortical tissue, we know that all types of aggression engage circuits that connect the brainstem, hypothalamus, limbic system, and forebrain (Adams, 2006; Davidson, Putnam, & Larson, 2000; Karli, 2006; Mattson, 2003; Nelson, & Trainor, 2007; Siegel, 2004; Victoroff, 2009). It is legitimate to speculate that, as adaptive advantages accrued to creatures with more and more encephalization and more and more proportional isocortical-to-allocortical tissue, the functions of evolutionarily earlier layers were incorporated into webs of connectivity that took advantage of the outcomes of the prior two billion years of selection. But evolution is never complete. That blending, for instance, of three layered allocortical limbic tissue with six-layered isocortical tissue to mediate motor actions in primates does not necessarily lead to perfect outcomes (e.g., Gardner & Cory, 2002; MacLean, 1990; Panksepp, 1998; Papez; 1937; Striedter, 2005). The thinner, older cortex may tend to direct one action in response to circumstances when the thicker, newer cortex would tend to direct another. When one is sideswiped in traffic and hurt by the screeching motor vehicle of a slightly callous, mildly attention disordered, drunken, unfamiliar teenaged boy, limbic and peri-limbic allocortex may quickly start organizing an aggressive behavioral response while prefrontal isocortex organizes restraint. Killing the teen would perhaps have been adaptive over most of the course of mammalian evolution. Not killing him might be more adaptive now that one is embedded in a large cooperative group with ferocious agreed-upon punishments for non-sanctioned aggression. Neuroscience may provide somewhat accurate, arguably deterministic explanations both for the offender's and the offended's behaviors. The question becomes, in social groups the individual members of which mutually profit by maintaining behaviors within a certain spectrum of sanctionability, when aggressive behaviors fall outside that spectrum, what is to be done?

That massive question is beyond the present domain of inquiry. This essay will introduce a much narrower issue: can the science of aggression help answer that human social and legal question, what is to be done? As will be discussed further, below, social evolution has proceeded rapidly even if human brain evolution proceeds slowly. Emerging from social evolution is a generally agreed upon notion of moral responsibility. But moral responsibility is (a) an abstract idea and (b) probably an instinct (e.g. Gazzaniga, 2005; Gruter, 1979, 1991; Mobbs, Lau, Jones, & Frith, 2007; Morse, 2004a,b; Murphy & Brown, 2007; Prinz, 2008; Roskies, 2008; Sie & Wouters, 2008; Walsh, 2000; Wigley, 2007; Wilson, 1993). It may well have neural correlates somewhat measurable by empirical methods (e.g., Hsu, Anen, & Quartz, 2008; Robertson et al., 2007; Takahashi et al., 2008; Young & Saxe, 2008a,b; Zahn et al., 2009), but it is not a purely objective neurological phenomenon. It is, among other things, a philosophical concept with deep roots in human intuition about others' minds.

The question is, when humans judge the moral responsibility of other humans—as one step toward answering the "what is to be done" question—might knowledge of the science of aggression be useful? Or, to narrow the issue even further and introduce the fact that written laws exist to address what is to be done, Morse and Hoffman (2007) put the modern question succinctly: " Should evolving ideas about the nature and causes of mental disorders and of behavior in general require changes in our settled views of blameworthiness?"

To outline the argument in chief: Aggression is a vital part of animal behavior. In social species, most aggression is sanctioned. Yet for social life to work, groups must punish those who commit non-sanctioned aggression. To maximize buy-in by cooperators, there must be some exceptions to the standard system of punishing non-sanctioned aggression. There is a biological instinctive basis for some universally agreed upon exceptions to punishment. The concepts of moral responsibility and *mens rea* evolved to formalize those biologically instinctive exceptions. But the idea of moral responsibility is rooted in a belief in the Kantian imperatives of both rationality and autonomy, a.k.a., free will—neither of which are amenable to scientific demonstration—and *mens rea* is based upon simplistic assumptions about how brains and minds work. As a result, the conventional understandings of moral responsibility and *mens rea* are suboptimal guides for a scientifically logical, biologically coherent system to excuse some non-sanctioned aggression. The neuroscience of aggression is making slow progress toward a better understanding of aggression. That science helps explain how given acts of non-sanctioned aggression fit or do not fit within the natural system of exceptions to punishment—a step toward the Kantian ideal of perfect (meaning exceptionless) rules. But the application of scientific views of aggression to law will be limited, because accepting their implications requires acknowledging elements of determinism, and humans have excellent biological reasons for resisting determinism and, instead, maintaining a strong faith in free will.

This commentary is merely an introduction—and an invitation—to a vibrant dialogue. To fully appreciate the potential implications of new findings in the science of aggression to the process of law and justice, it would be necessary to tackle the profound philosophical issues that have bedeviled the field for millennia. At the heart of that suite of tough questions is whether or not humans have free will. The free will question underpins the gamut of equally tough questions faced by those attempting to enhance justice by better understanding of human nature. What is moral responsibility? What deviations from norms of brain or mental function alter or constitute excuses for deviations from either (a) conative moral reasoning or (b) appreciation of moral conventions or (c) capacity to comport oneself with those conventions? Do variations from the typical course of human life (say, 6 min of birth hypoxia, or being a victim of child rape, or suffering intermittent epileptic seizures, or bearing a gene mutation significantly affecting neurotransmission, or recovering from a massive frontal lobe traumatic brain injury) represent exculpatory or mitigating circumstances for non-sanctioned aggression? At precisely what degree of brain atypicality, demonstrated with what degree of confidence, shall humans grant one another exceptions to the instinct for punishing non-sanctioned aggression?

This introductory essay cannot hope to review, in depth, the history of the interplay between the scientific understanding of aggression, of neural function, and of the law, far less resolve the question of free will. Indeed, one purpose of this essay will be to explain why resolution of that question to the satisfaction of all persons is biologically improbable. However, I will briefly review several key points in the history of this dialogue, and will offer a framework that may prove useful in contemplating how the science of aggression can and cannot contribute to law and justice.

## 1. The evolution of restraints on aggression

Before plunging into the thorny scientific–legal questions of responsibility, *mens rea*, and the neuroscience of aggression, it is important to clarify that one is ultimately discussing the evolutionary purpose of law.

Accepting that no single definition of aggression will achieve universal acclamation, one nonetheless identifies behaviors that do or do not satisfy an instinctive understanding of what it means to aggress. Among humans, physical contention is generally recognized as aggressive. The peak of physical contention and unequivocal interpersonal assaultiveness occurs at about age 2.5 and is more frequently observed among boys than girls (e.g., Loeber & Hay, 1997; Olweus, 1979; Pettit, 1997). This behavior is probably an adaptation that prepares for more life-and-death contention later in life—as does the physical contention of sport. Parental discipline almost universally involves some physical contention and is also highly adaptive. Physical contention directed at animals, either in self-defense or to acquire calories and protein is also highly adaptive. And as groups form larger

than the immediate family, physically aggressive behaviors are directed at non-kin are valuable to police the mandates of social cooperation. And whenever groups of identity form that cooperatively share resources, physical contention with other groups over those resources may occur. All of these types of aggression—toddlerhood, discipline, sport, defense, hunting, policing, and military action— are normal, natural, and socially sanctioned. It is entirely possible that the overwhelming majority of instances of human aggression are sanctioned. The essence of sanctioned aggression is that it fits within the evolutionary framework of both individual adaptation and group cooperation. This leads to the question of the evolution of cooperation.

For social organisms and species to survive the winnowing process of natural selection, such organisms have been shown—via quantitative evolutionary biology—to need to cooperate (e.g., Axelrod & Hamilton, 1981; Dugatkin, 2002). Group cooperation requires unwritten but commonly embraced rules. The simplest case of such cooperative rules is kin selective behavior, or actions to promote inclusive fitness by assisting those who share the largest number of genes (Hamilton, 1964). At the next level of complexity is reciprocal altruism, or the shared understanding that group members, even non-kin, will help another on the assumption that the favor will be returned (Dugatkin, 2006; Mesterton-Gibbons & Dugatkin, 1992; Trivers, 1971). A *strong* form of reciprocal altruism allows that group members are expected to return the favor not just to the particular individual who previously helped, but to other recognized ingroup members, related or not (e.g., Gintis, 2000; Panchanathan & Boyd, 2004).

For strong reciprocal altruism to work, there must have evolved rewards for pro-social and punishments for anti-social behavior. In part, that policing must become internal. Internalized values—probably mediated by genes and expressed depending on early environmental factors—lead to the emotions of shame and guilt for violations of the rules of society. To discuss the fascinating cognitive neuroscientific evidence for such internal policing is beyond the scope of the present essay, although readers will intuit that something must be wrong with the internal policing of psychopaths. For the purposes of explaining the evolution of laws (or, more precisely, the evolution of genes for nervous systems that, via co-evolution of genes and culture, originated laws), it is more pertinent to discuss external policing of reciprocal altruism—the *sine qua non* of survival for social species. Different species police reciprocal altruism differently. Bee police, for instance, destroy larvae of rogue female workers that try to cheat by laying eggs in the queen's honeycomb. The rules of punishment for behavioral deviations from the social order are similar throughout the animal kingdom. And, standing as a natural barrier to the idealistic ambitions of scholars who urge humans to apply reason to transcend retributive justice (e.g., Greene & Cohen, 2004; Slobogin, 2005), punishment has been shown to be essential for the survival of social groups (e.g., Boyd & Richerson, 1992; Clutton-Brock & Parker, 1995; Rockenbach & Milinski, 2006; Seymour, Singer, & Dolan, 2007. Also see Buckholtz et al., 2008; Hoffman & Goldsmith, 2004). Socially sanctioned aggression is rewarded, or, at least, tolerated. Non-sanctioned aggression must be punished. Long before there was law, or writing, or even speech, social species necessarily evolved universal rules, within an evolutionarily acceptable range of genetic and cultural intergroup variation, to punish non-sanctioned aggression.

The evolved restraints on aggressive behavior in social species, with rewards for most aggression and punishments for some aggression, are applied somewhat differently in different genera. In a hard-wired species with a nervous system functioning a little above the computational level of the knee-jerk reflex, the rules for distinguishing sanctioned from non-sanctioned behaviors need to be stereotypical. Social insects like bees, for example, behave as if hard-wired to reward and punish without exceptions. The rules are Kantianly perfect, in the sense that they are exceptionless. You behave this way, the group will respond that way. But other species have evolved nervous systems that can make rewards and punishments conditional. With sufficiently complex computational capacity, those species might be equipped to consider exceptions to knee jerk justice. An instance of aggression might fall into the non-sanctioned class, yet additional factors might be used to evaluate the evolutionarily adaptive value of applying the conventional punishment. That is, for very good reasons, some species may have the capacity to make strategic exceptions to knee jerk rules. Humans appear to have such complex nervous systems.

## 2. Some species behave as if subjectively self-conscious, believe themselves to have free will, and believe themselves to possess rationality

Bees may or may not be self-conscious. But humans behave, to all appearances, as if they are self-conscious. For whatever reason (and the reason remains hotly debated) via the simple process of selection of fitness-promoting adaptations, genes seem to have evolved that, after epigenetic interactions with the environment, mediate the development and operation of human brains with subjectivity—the very useful, arguably illusory, and apparently material experience of self consciousness (e.g., Bering & Shackleford, 2004; Kahane & Savulescu, 2009; Wigley, 2007). Genes also orchestrate neuronal ontogeny leading to a subjective sense of agency, or a belief in one's own free will (Ainslie, 2001; Bok, 2007; Frankfurt, 1971; Morris, 2007; Morse, 2007a; Murphy, 2007; Rakos, Laurene, Skala, & Slane, 2008; Roskies, 2008; Sie & Wouters, 2008; Taylor & Dennet, 2001). The adaptive value of a subjective sense of free will is apparent. As Allison (1997), Pereboom (2001), Slote (1990), Wegner (2002), and others have argued, despite the rational objections to the existence of free will as anything but an illusory perception, in the presence of subjective sense of self consciousness, organisms would be hard put to take action without believing that it is they who are taking that action.

The questions of the material versus non-material basis of consciousness and free will are challenging and hoary. The participants in the debate tend to fall into the camps of determinists (who understand all behavior to follow material rules) and non-determinists (who don't) and compatibilists (who understand material determinism to be compatible with free will) and non-compatibilists (who don't) (see, e.g., Kane, 1996). Admitting the extreme complexity of the philosophical debate and the limitations of the present introductory discussion, at this point I must confess to two non-universal opinions: First, I have yet to encounter a scientifically coherent account for a non-material basis of consciousness. Second, I have yet to encounter a coherent physical explanation for free will. For example, Dennett's (1995, 2003) argument for "life worlds" in which deterministic rules lead to variable outcomes supposedly supports free will. It doesn't. It merely supports stochastic elements in mechanical determinism. Much has been written about Libet's famous psychophysiological experiments demonstrating a delay between the brain's readiness potential and the occurrence of willed movement (e.g., Libet, Gleason, Wright, & Pearl, 1983; Libet, 1993, 1999). These experiments are usually interpreted as evidence that brains make decisions before minds do, contrary to the notion of free will. Gazzaniga (2005, 2007) has argued that the 50–100 ms interval between the onset of the readiness potential and the hand movement in Libet's experiments is an opportunity for "free won't"—an opportunity for a self-conscious agent to voluntarily restrain an impulse—and therefore that a type of free will exists. With respect, this is fallacious. The presence of an available window of time to exert constraint is not a logical argument that constraint occurs.

Absent a scientifically coherent account for non-material consciousness or material free will, to hold that consciousness is something apart from an emergent property of matter, or to hold that material organisms possess free will that permits agency independent of the laws of matter means that one is advancing a

spiritual belief, not science. But, as will be elaborated below, these beliefs are at the heart of the law.

Self-conscious organisms also appear to regard themselves as making self-interested decisions. In the language of economists, humans believe themselves to be maximizers of utility under uncertainty. In the language of some philosophers, humans believe themselves to be rational. Again, this brief introductory essay cannot comprehensively address the problems with the claim of human rationality. Suffice it to say that abundant empirical evidence demonstrates that violations of rationality are normal, routine, and expected in real human decisions. Among the factors contributing to such violations are framing effects (De Martino, Kumaran, Seymour, & Dolan, 2006), choice blindness (Johansson, Hall, Sikstrom, & Olsson, 2005), and the fact that, utility aside, people are biased to favor their own decisions (e.g., Egan, Santos, & Bloom, 2007). Contrary to the prediction of economics, ambiguity about probabilities does not reliably affect human brain decisions (Hsu, Bhatt, Adolphs, Tranel, & Camerer, 2005). In fact, pre-conscious mechanisms may bias perceptions of sensory stimuli to fit with half-made decisions, (Summerfield et al., 2006). This may help explain the fact that emotion also obviously affects decision-making (e.g., Lerner & Tiedens, 2006; Naqvia, Shiv, & Bechara, 2006). And it is perhaps no surprise that sexual arousal has been shown to bias decisions (Ariely & Loewentsein, 2006). Especially important for jurisprudence, humans' irrational enthusiasm for punishment often leads to choosing punishment over self-interest (e.g., Herrmann, Thöni, & Gächter, 2008). While it may be highly adaptive for humans to regard themselves as rational agents arriving at their own decisions, some evidence suggests not only that decisions are often irrational but that conscious "decision-making" is epiphenomenal, a post-hoc rationalization of brain-algorithm-generated outputs (e.g. Bechara, Damasio, Tranel, & Damasio, 1997; Libet et al., 1983).

Some thinkers have excused the violations of rationality as evidence of so-called "bounded rationality," proposing that one can only devote so much brain effort to thinking things through and that energy limitations mandate cognitive short cuts (e.g., Simon, 1997). But the theory of bounded rationality is not consistent with the evidence that, independent of the time available or the life-and-death importance of the decision, humans simply and routinely make irrational choices. Cerebral energy conservation is not the rate-limit step in rational decision-making. Humans' normal, healthy irrationality is better explained by variable compliance with evolutionary fitness seeking rules that are insufficiently explained by classical macro-economics.

Thus, humans perceive subjective self-consciousness, free will, and rationality. Each of these three probably has potent fitness implications. Aggressive behaviors are usually discussed within the framework of these assumptions. But it is difficult to prove that first is anything but a physical phenomenon, that the second exists, and that the third is true. Knowing how vital these beliefs are for humans helps to explain the historical evolution and inherent fallacies of *mens rea*.

## 3. A theory of mind enhances the social behavior of subjectively self-conscious organisms

In order to account fully for some aspects of aggression, and to account at all for the evolved instinct for just responses to non-sanctioned aggression, it is insufficient to have subjective self-consciousness, subjective free will, and (unjustifiable) faith in self-rationality. Another necessary element is a theory of mind. A theory of mind is the belief that other members of one's species are also subjectively self-conscious (Astington, Harris, & Olson, 2001; Baron-Cohen, 1995; Baron-Cohen, Tager-Flusberg, & Cohen, 2000; Premack & Woodruff, 1978). In fact, it might be proposed that the principal adaptive virtue of self-consciousness is that it permits one human to make informed guesses about the thinking of another. The better one can guess the internal and invisible mental operations of another person, the better one can predict that other's future behaviors in response to contingencies. Evidence suggests that the mirror–neuron system, especially at the temporo-parietal junction, evolved to facilitate the theory of mind by, in essence, activating similar clusters of brain cells in the actor and the observer (e.g., Carruthers & Smith, 1996; Gallagher & Frith, 2003; Pelphrey, Morris, & McCarthy, 2004; Premack & Woodruff, 1978; Rizzolatti & Craighero, 2004; Saxe & Kanwisher, 2003, Saxe & Powell, 2006). Genes support brains with a theory of mind, or a sensitivity to the likelihood that other organisms also subjectively experience (a) consciousness, (b) autonomy or free will and (b) a faith in self-rationality similar to one's own.

Much has been written about the evolution and processes involved in a human theory of mind, but two facts of critical relevance to the evolution of law are often neglected. First, a theory of mind assumes similar minds. But everyone is different. The theory of mind evolved to anticipate a fairly narrow range of differences in mentation from oneself. Adult humans cannot perfectly estimate another's mentation since the man before them may or may not have an especially atypical mind. People can accommodate universal expectations from their own mentation such as the mind of an infant or a sleeper. People are shocked and baffled by deviant thinking such as mental retardation, dementia, or psychosis.

Second, and a closely related problem, since humans imagine themselves to be rational, they expect that others will also be rational.

A theory of mind helps to explain some types of non-sanctioned aggression. The bank robber guesses that a witness may turn him in and may kill based upon this guess about the witness's intentionality. Some psychopaths imagine all others to share their outlook and may kill a person to whom the killer incorrectly attributes remorseless self-interest. The paranoid person attributes malice where there is none and may kill the person whose mind he has incorrectly estimated.

But a theory of mind is even more vitally linked to the evolution of justice. Humans generally assume that perpetrators of non-sanctioned aggression have minds similar to their own. Blameworthiness is judged based largely upon that assumption. Punishment for behavior that threatens the inclusive fitness of the average member of the ingroup derives from the best guess that the perpetrator acted for pretty much the same motives as the judge might have—rational self interest, under perfect self control, contrary to the rules of strong reciprocal altruism. Punishment is also based upon the judge's guess that the perpetrator's mind is similar enough to his own that the punishment will be deterrent. Triers of fact in tribal councils or supreme courts begin their deliberations about blameworthiness and punishment with this assumption of mental similarity. The reader can anticipate the result given the insupportability of the two core assumption of theory of mind, mental similarity and rationality: inflexible rules for punishment that try to cover all minds will produce outcomes that do not serve the underlying purpose of the instinct for justice, to maintain social cooperation.

## 4. The evolution of morality, justice, and exceptions

Humans, therefore, are the product of more than three billion years of selection that endowed them with subjective self-consciousness, subjective free will, biased faith in their own rationality, and some ability to anticipate the behavior of others based upon the imperfect assumptions of the theory of mind. All of these elements are part of any comprehensive theory of the evolution of morality. Again, it is beyond the scope of this essay to provide a full account of the evidence for an evolved, subconscious, neurally mediated instinct for distinguishing between right and wrong and just and unjust. Many excellent empirical studies and discussions are available (e.g., Casebeer, 2003; de Waal, 1996; de Waal, Macedo, & Ober, 2006; Goodenough, 2001; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Haushofer & Fehr, 2008; Katz, 2000; Prinz, 2008; Ridely, 1996; Singer, 2007; Wright, 1994). For the limited purposes of examining how the science of aggression can and cannot advance the administration of justice, a very brief summary should be sufficient.

In order to maintain the fitness-benefiting operation of human social groups, long before there was writing, rough rules of rewards and punishments for sanctioned versus non-sanctioned aggression must have become part of the adult mental repertoire. Humans intuit that other humans share their own basic understanding of the strict rules of reciprocal altruism and also "know" the rewards and punishments one should expect for compliance with or violations of reciprocal altruism. This basic understanding philosophically conceptualized as "responsibility" and culturally instantiated in the so-called golden rule. Primatologists observe apparent conformance with the golden rule in multiple species of ape (e.g., de Waal, 1996; Sober and Wilson, 1998; Wilson, 1993). Therefore, the golden rule has plausibly been part of the intuitive moral psychology of hominids for millions of years prior to the split, about 6 million years ago, between the lineage that led to chimpanzees and that which led to man.

This instinct for justice begins with the assumption of mental similarity (see, e.g., Goodenough, 2004; Jones, 2000). But since some instances of mental difference are familiar and obvious, the instinct for justice also accommodates *exceptions* to the rules. It is accepted throughout the world that certain classes of persons should be regarded as less culpable than others for the same offense. The 2-year-old trigger-puller is universally regarded as less blameworthy than the normal adult. So is the congenitally profoundly retarded person. So is the sleeper who, twitching in his unconsciousness, rolls off the bed and crushes an infant to death. The reasons for these universal intuitive exceptions to full culpability have to do with the universal familiarity of humans with the mental otherness of childhood, the unconsciousness of sleep, and, to a lesser extent and accompanied by more fear and uncertainty, with the otherness of mental retardation and senile dementia. Note that the person with developmental delay is typically assigned a lower age equivalent, simplifying his condition for others by equating him with a younger child. Similarly, the Alzheimer's afflicted person is also often referred to as "childlike," and the person aggressing in the frank unconsciousness of a seizure (or in sleep disorder) is often explained by reference to normal sleep. Thus, the genetically coded theory of mind comfortably generalizes the universal exceptions of childhood and sleep/unconsciousness to a few other special cases.

## 5. Western Law formalizes exceptions to the rules of punishment for non-sanctioned aggression: The evolution of law and the problems with *mens rea*

The genetically encoded, epigenetically expressed golden rule underpins laws (see, e.g., Cosmides & Tooby, 2006). At some point in the misty past, a nascent concept of criminal responsibility took hold within the framework of a category of abstract thought now called legal reasoning. Individuals who violated the golden rule—a popular restatement of strong reciprocal altruism—could be regarded as "responsible" since others intuit that the violator had subjectivity and free will, more or less identical to their own. Laws evolved to formalize and verbalize the instinct for justice that grants humans immediate access to neural systems that evolved to guess why another person did something, whether or not he was "responsible," and what should be done to mitigate the harm to social cooperation caused by such a person (e.g., Hinde, 2004). And laws have struggled for millennia to formalize and verbalize the exceptions. This raises the question of *mens rea*.

*Mens rea* in Western law probably derives from the Latin postulate, "*actus non facit reum nisi mens sit rea*" or *an act does not make for guilt unless the mind is guilty*. Some scholars attribute this postulate to Augustine's writings of 579 C.E., and assert that Anglo-Saxon legal thinking has embraced this principle since before 1100 C.E. (e.g., Raymond, 1936). Hence, for about a millennium, assumptions implicit in Western law have included that (a) humans who perform acts of non-sanctioned violence will usually experience a particular type of mental change called the guilty mind, and (b) if, for any reason, their minds do not experience this change, then they are not criminally responsible. Both assumptions are debatable. A two year old who mistakenly shoots his father may indeed feel guilty but is not held responsible. A person who delusionally believes he has committed an unsolved murder and gives a false confession has a guilty mind but is not responsible. A psychopath who feels no guilt is held responsible. The fact that it is so easy to find counter-examples to the common understanding of *mens rea* hints at the weakness of this concept.

*Mens rea* evolved (as nicely outlined by Phillips and Woodman, 2007) when the English legal scholar popularly known as Bracton (c. 1210–1268) laid out the excuses by which children and the insane should not be held liable: the first lack malice and the second lack reason (see Sayre, 1932). Blackstone codified these notions in the 18th century, stating, "So that to constitute a crime against human law, there must be, first, a vicious will; and, secondly, an unlawful act consequent upon such vicious will" (cited in Swanson, 2002). Blackstone (1769) specifically tied blameworthiness to free will, stating that it is a deficiency of will makes the lunatic unable to distinguish between right and wrong. Yet the strict application of the "vicious will" standard led to some undesirable results: the highwayman who fires a warning shot that unintentionally kills a princess is blameless. The M'Naughten case of 1843 revised Blackstone's standard by describing two excuses from blameworthiness for the mentally ill: at the time of the offense (1) the offender did not know the nature and quality of his actions, or (2) did not know that what he was doing was wrong.

The remarkable vagueness and immeasurability of these two exceptions is immediately apparent. Experts cannot know whether, or at what precise moment in the course of a past action, the perpetrator knew the nature and quality of his actions and categorized them either according to morality (right or wrong) or convention (legal or illegal) (see, e.g., Denno, 2002). And the M'Naughten standard, which emphasizes the guilty mind, weakly addresses the issue of diminished capacity and completely fails to address the possibility of irresistible impulse (e.g., Carter & Hall, 2007; Morse, 2002, 2007b; Phillips & Woodman, 2007; Vincent, 2008). Indeed, as Barratt and Felthous (2003) have cautioned, despite considerable progress in understanding impulsivity, the concept of *mens rea* and the major iterations of the insanity defense (including the hybrid "extreme emotional disturbance" defense) are poor ammunition against the natural resistance to excepting explosive but temporary brain states in otherwise sane persons.

Since M'Naughen, various equally problematic formulations have emerged (e.g., Elliott, 1996). Nonetheless, I would argue that the very adaptive instinct for justice and the universal recognition of exceptions to punishments for some classes of non-sanctioned aggression have driven a massive psychic enterprise (and generated a massive scholarly literature) all in search of a perfect rule that will formalize and verbalize the instinct to except some perpetrators from the usual punishments.

Leaping ahead chronologically, the U.S. Supreme Court has enshrined various aspects of the intuited exceptions to moral responsibility. In Ford v. Wainwright (1986), the court held that it is constitutionally prohibited to execute an insane offender, and in Panetti v. Quarterman (2007), the court clarified that even a person found mentally competent to stand trial may qualify for this exception. In Thompson v. Oklahoma (1988), in Stanford v. Kentucky (1989), and in Roper v. Simmons (2005) the court forbade capital punishment for the less than fully responsible minor person. In Penry v. Lynaugh (1989) and in Atkins v. Virginia (2002), the court forbade the execution of the mentally retarded. Setting aside the much-debated question of the morality of capital punishment (e.g., Bedeau, 1987; Berns, 1979; Torr & Egendorf, 2000), the essence of each of these decisions was a judgment that their wordings would formalize and verbalize the instinct for justice (see, e.g., Jones, 2006; Mobbs, Lau, Jones, and Frith, 2007). There are natural exceptions to the usual understanding of responsibility. For all the considered verbiage of these decisions, from the point of view of the evolutionary psychology of law, one might simply have said, "These cases fall within the exceptions to the

attribution of full responsibility universally recognized as a result of natural selection."

## 6. Natural limits to the contribution of the neuroscience of aggression to the administration of justice

Some types of non-sanctioned aggression are readily judged to fit within the natural exceptions to punishments, and laws have formalized these exceptions. But experts in neurobehavior and attorneys know that controversy persists. Death penalty appeals are argued, often for decades, because it remains challenging to apply the limited available rules to the infinitude of human circumstances. It is legitimate to hope that advances in the neuroscience of aggression, such as those reported in the present Special Issue of the International Journal of Law and Psychiatry (IJLP), will have the side effect of modestly enhancing justice.

Neuroscience will eventually provide highly valid and reliable determination of truth telling (Appelbaum, 2007; Harada et al., 2009; Illes, 2007; Keckler, 2006; Spence et al., 2006; Van Hooff, 2008). This by itself is a potentially revolutionary occurrence, since it may divert the usual assignment of the role of trier of fact to a machine. But neuro-lie-detection is not directly pertinent to the biology of aggression. Setting aside expected advances in lie detection, there are a number of ways in which, theoretically, aggression science might be useful to the law (e.g., Goodenough, 1998; O'Hara, 2004; Morse, 2006; Prehn et al., 2008; Stefánsson, 2007; Tancredi, 2005; Wolf, 2008). Neurosciences will evolve over the next 100 years to the point at which judges and juries no longer need to guess how closely the offender's mind matches their own or matches the typical adult mind that was anticipated by the framers when various rewards and punishments were devised in a given society. A social or neuroscientist might help measure:

a. Capacity to understand socially conventional "right and wrong," and the boundaries between sanctioned versus non-sanctioned aggression
b. Capacity for conventional moral "reasoning"
c. Capacity to distinguish real (consensual subjective reality) versus unreal
d. Capacity to reasonably assess threats
e. Capacity to comport oneself with society's expectations for restraint of aggression, especially impulse control under stress
f. Capacity to resist urges such as substance addiction
g. Biological sources of variation from normal capacity, e.g., genetic, epigenetic, neurochemical, structural or functional
h. Statistical likelihood of again aggressing in a non-sanctioned way within a given timeframe,
i. Likelihood of being deterred from aggressing by receiving conventional punishments
j. Neural susceptibility to rehabilitation, etc.

Soon, therefore, the science of aggression and closely related forensic neurosciences will tell us, with known measures of validity and reliability, that murderer A has a:

- 99.9% likelihood of being a liar when he denies having pulled the trigger,
- 92% likelihood of being a liar when he denies having *intended* to pull the trigger,
- 58% likelihood of being a liar when he denies *wishing the victim dead* as a result of pulling the trigger,
- 78% of the normal adult human capacity for restraint of aggression under moderate stress,
- 44% of the normal adult human capacity for restraint of aggression under high stress,
- 63% of the normal adult human capacity to distinguish real from non-real threat in response to standardized stimuli,
- 50% of the normal adult human mirror neuron response in theory of mind tasks,
- 190% of the normal adult human level of paranoia in response to standardized exposures,
- 378% of the normal adult human genetically-based, epigenetically expressed neurobiological vulnerability to stimulant addiction, and
- 64% of the normal adult human capacity to distinguish "right from wrong" when confronted with hypothetical moral choices about aggressive versus non-aggressive behavioral options.

The question is whether, how, and to what degree, these knowledge enhancements will or should impact the administration of justice. Some scholars are optimistic. Greene and Cohen (2004), for example, opined that

> "…neuroscience will probably have a transformative effect on the law, despite the fact that existing legal doctrine can, in principle, accommodate whatever neuroscience will tell us. New neuroscience will change the law, not by undermining its current assumptions, but by transforming people's moral intuitions about free will and responsibility." (p. 1775).

Initiatives such as those of the U.S. Social Science Research Council and the Dana Foundation promise advancement in the achievement of just outcomes by enlightenment about the neuroscience of morality. Yet, as one probes the arguments of the participants in debates about consciousness, free will, decision-making, *mens rea*, and the neuro-philosophy of moral responsibility, certain terms recur.

Debaters refer to one another as hard or soft determinists (respectively, those who believe that all is determined by physical laws versus those who believe that, despite the role of cause and effect, humans have agency), compatabilists (those who believe that determinism and free will are somehow compatible), rationalists (those who believe that humans are rational optimizers of choice under uncertainty), dualists (those who believe that body and mind are separable), internalists (those who believe that good action is intrinsic), sentimentalists (those who believe that our intrinsic moral judgments compel us toward good action) , consequentialists, etc. The use of the "ist" label is revealing. It demonstrates that even highly educated, intellectually able, knowledgeable persons who might be regarded as objective end up being categorized by one another with terminology reserved, in philosophy and anthropology both, for true believers. There may or may not be one correct, biologically valid and philosophically coherent explanation for human moral responsibility. But the profusion and persistence of "ists" and "isms," 2000 years into the debate, strongly suggests that instinct, opinion, and belief dominates the discussion. This means that, no matter how persuasive a bit of scientific discovery might be regarding brain mediation of aggression, one cannot expect many minds to be changed by it in the near future. Determinists, who feel in their bones that the material basis of the human mind is absolutely and self-evidently incompatible with free will cannot understand the soft-minded behavior of those soft-determinists who struggle mightily to come up with a scenario in which materialism and free will are magically compatible. Soft-determinists, passionately certain that there is free will, cannot abide the materialism of the strict determinists who, they suppose, are simply resisting the undeniability of human freedom for the sake of a sterile vision of the universe unwinding like a clock.

The dominance of "isms" in these vital debates also hints at the reason that these differences are not likely to be concluded with agreement. Beliefs are sacred in the mind and not amenable to debate (e.g., Atran, 2002; Boyer, 2001; Hoffer, 1951; Newberg, D'Aquill, & Rause, 2001). Beliefs about one's own mind and about other's minds are also valuable products of evolutionary adaptation. As social species with subjective self-awareness arrived on earth, those bearing gene variants that permitted better prediction of other's behaviors flourished. Having a theory of mind is not a cultural trait but a universal feature of humans. But different individuals (and perhaps

different cultures) may develop brains that process the stimuli relevant to theory of mind in different ways. Among the varieties of human variety, one is the difference between people in how they explain human nature to themselves. One fellow will be certain that the murderer acted by conscious decision-making based upon free will, and resolve to put that murderer away for 25 years. Another will be certain that the murderer became dangerous because birth hypoxia and childhood trauma interacted with an unfortunate monoamine oxidase gene polymorphism to make him prone to aggression and a dopamine receptor gene polymorphism made him vulnerable to addiction, and then an adult orbitofrontal lobe head trauma gutted his self-control such that he rashly pulled a trigger at a stressful moment when he was intoxicated—and resolve to put him away for 20 years. The particular way one rationalizes one's in-born theory of mind may not be the major determinant of decisions in day-to-day life or law. Humans dispense justice so that cooperation is preserved. Even in dealing with extraordinary threats to familial and group security, such as the recidivist rapist and serial murderer, judgment is less based upon which rationalization one has embraced than upon pragmatic instincts about what judicial action might satisfy emotional needs and enhance personal and in-group survival. Outcomes are bound to hew closely to what has proved useful, on average, throughout the history of the species, for making social life work.

Since the different positions in the debate about the biological basis of aggression and moral responsibility may themselves be rooted in inescapable biology (and in fact some evidence suggests that cognitive traits such conservatism and punitiveness may be influenced by gene variation (e.g., Alford & Hibbing, 2007; Koenig & Bouchard, 2006), the science of human aggression can expect to make limited inroads toward advancing the fairness of laws. Mental health experts and jurists who believe in free will are unlikely to read a scholarly article that changes their minds, anymore than adult believers in God will be convinced by a tract that there is no god. More to the point, if the subjective perception of free will is so valuable an adaptation, then theses on its illusory nature will be as readily embraced as those telling the reader that he is not conscious. The manifest logical virtues of determinism aside, it goes against the rough grain of human nature to tolerate more than a little wiggle room in the instincts for justice.

The realization that humans are naturally constrained, by virtue of their biology, in their capacity to absorb scientific information about aggression and use it to modify their instincts for justice should not be tarred with another belief label: fatalism. This realization is merely the exciting discovery and acceptance of human limits, or what Konner called, in *The Tangled Wing* (*2003*), the "biological constraints on the human spirit." Tancredi (2005, p. 81) explained these human limits succinctly: "…nature has the dominant role in explaining the range and variability of the mind." Yes, civilization advances. Cultures can, within limits, adopt more nuanced ways to formalize and verbalize the instinctive rules that maintain strong reciprocal altruism. Cultures will continue to struggle to find closer-to-optimum ways to generalize from the natural exceptions—the reduced responsibility of the unconscious person, the child, or the person who is profoundly alienated from consensual reality. But, absent biological evolution and superhuman brains, the limits will stand. Triers of fact, equipped with very human brains, can only conceptualize a given case of unacceptable behavior within their instinctive, universal framework of reduced responsibility. Triers of fact can only be moved so far toward accepting that a perpetrator of a ghastly act should logically be granted an exception that evolved, in the mind and in the law, to fit a more familiar case. For this reason, it is unrealistically sanguine to predict that advances in the neuroscience of aggression will "transform" the law. Rather than to pour immense energy into the doomed exercise of trying to bring judges and jurors around to opinions that will forever grate on the human spirit, it seems wiser to acknowledge and study these human constraints—and to serve the cause of justice by developing neurobehaviorally sophisticated, philosophically logical, and intuitively persuasive ways to explain how a given case of non-sanctioned aggression fits or does not fit with the deep purpose of natural exceptions.

In summary, it is not clear how advances in the science of aggression will, or should, change the law. The reason it is not clear relates to the essential disconnect between the goals of science and those of law (e.g., Aharoni, Funk, Sinnott-Armstrong, & Gazzaniga, 2008; Jones, 2004; Waldbauer & Gazzaniga, 2001) and, in particular, the yet-to-be-widely-adopted evolutionary framework (e.g., Stake, 2001; Terry, 2002; Walsh, 2000). Science sees behavior as that near infinitude of possibilities arising from material brain activity. Law formalizes unconscious intuitions about other's minds—including intentionality, free will, consciousness, self-control, and likely response to sanctions—that evolved to assist in arriving at socially useful systems of rewards and punishments. Laws demand simple formulae permitting small adjustments in primitive intuitions about other's minds in order to reach biologically simplistic conclusions about responsibility or non-responsibility, whereas scientists must struggle when asked to contribute information of value to the gross oversimplification of human behavior implied by the guilt-versus-innocence framework. But we try.

## Acknowledgements

This work was supported, in part, by a grant from the Freya Foundation for Brain, Behavior, and Society. The author wishes to acknowledge the extremely valuable executive contributions to the Special Issue on Aggression of Nina Marie Fusco, Associate Executive Director, International Academy of Law and Mental Health, and of Janice Adelman, M.Sc., Department of Psychology, Claremont Graduate Universities.

## References

Adams, D. B. (2006). Brain mechanisms of aggressive behavior: An updated review. *Neuroscience and Biobehavioral Reviews*, *30*, 304—308.
Aharoni, E., Funk, C., Sinnott-Armstrong, W., & Gazzaniga, M. (2008). Can neurological evidence help courts assess criminal responsibility? Lessons from law and neuroscience. *Annals of the New York Academy of Sciences*, *124*, 145—160.
Ainslie, G. (2001). *Breakdown of will.* Cambridge, UK: Cambridge University Press.
Alford, J. R., & Hibbing, J. R. (2007). Personal, interpersonal, and political temperaments. *Annals of the American Academy of Political and Social Science*, *614*, 196—212.
Allison, H. A. (1997). We can only act under the idea of freedom. *Proceedings of the American Philosophical Society*, *71*, 39—50.
Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, *53*, 27—51.
Appelbaum, P. S. (2007). Law & psychiatry: The new lie detectors: Neuroscience, deception, and the courts. *Psychiatric Services*, *58*, 460—462.
Ariely, D., & Loewentsein, G. (2006). The heat of the moment: The effect of sexual arousal on sexual decision making. *Journal of Behavioral Decision Making*, *19*, 87—98.
Astington, J., Harris, P. L., & Olson, D. R. E. (Eds.). (2001). *Developing theories of mind.* New York: Cambridge University Press.
Atkins v. Virginia, 536 U.S. 304 (2002).
Atran, S. (2002). *In Gods we trust: The evolutionary landscape of religion.* New York: Oxford University Press.
Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390—1396.
Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* Cambridge, MA: MIT Press.
Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. (Eds.). (2000). *Understanding other minds: Perspectives from developmental cognitive neuroscience.* Oxford: Oxford University Press.
Barratt, E. S., & Felthous, A. R. (2003). Impulsive versus premeditated aggression: Implications for *mens rea* decisions. *Behavioral Sciences and the Law*, *21*, 619—630.
Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*, 1293—1294.
Bedeau, H. A. (1987). *Death is different: Studies in the morality, law, and politics of capital punishment.* Boston: Northeastern University Press.
Bering, J. M., & Shackleford, T. K. (2004). The causal role of consciousness: A conceptual addendum to human evolutionary psychology. *Review of General Psychology*, *8*, 227—248.
Berns, W. (1979). *For capital punishment: Crime and the morality of the death penalty.* New York: Basic Books.
Blackstone, W. Commentaries on the laws of England 24 (1769/2002). Cited by Justice Scalia in his dissent in Atkins v. Virginia, 536 U.S., 304, 340, 122 S.Ct. 2242, 2260, 153, L.Ed. 2d 335 (2002).
Bok, H. (2007). The implications of advances in neuroscience for freedom of the will. *Neurotherapeutics*, *4*, 555—559.

Boyd, R., & Richerson, P. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, *13*, 171—195.
Boyer, P. (2001). *Religion explained: The evolutionary origins of religious thought.* New York: Basic Books.
Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., et al. (2008). The neural correlates of third-party punishment. *Neuron*, *60*, 930—940.
Bushman, B. J., & Anderson, C. A. (2001). Is it time to pull the plug on hostile versus instrumental aggression dichotomy? *Psychological Review*, *108*, 273—279.
Butler, A. B., & Hodos, W. (2005). *Comparative vertebrate neuroanatomy: Evolution and adaptation*, 2nd Ed. Hoboken: John Wiley & Sons.
Carruthers, P., & Smith, P. K. (Eds.). (1996). *Theories of theories of mind.* Cambridge: Cambridge University Press.
Carter, A., & Hall, W. (2007). The social implications of neurobiological explanations of resistible compulsions. *The American Journal of Bioethics*, *7*, 15—17.
Casebeer, W. D. (2003). Moral cognition and its neural constituents. *Nature Reviews. Neuroscience*, *4*, 840—846.
Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. *Nature*, *373*, 209—216.
Cosmides, L., & Tooby, J. (2006). Evolutionary psychology, moral heuristics, and the law. In G. Gigerenzer & C. Engel (Eds.), Heuristics and the law. Dahlem workshop reports. (pp. 175–205). Cambridge, MA, US: MIT Press; Berlin, Germany: Dahlem University Press.
Davidson, R. J., Putnam, K. M., & Larson, C. L. (2000). Dysfunction in the neural circuitry of emotion regulation—a possible prelude to violence. *Science*, *289*, 591—594.
De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science*, *313*, 684—687.
Dennett, D. C. (1995). *Darwin's dangerous idea: Evolution and the meanings of life.* New York: Simon & Schuster.
Dennett, D. C. (2003). *Freedom evolves.* New York: Viking.
Denno, D. W. (2002). Crime and consciousness: Science and involuntary acts. *Minnesota Law Review*, *87*, 269—399.
de Waal, F. B. M. (1996). *Good natured: The origins of right and wrong in humans and other animals.* Cambridge, MA: Harvard University Press.
de Waal, F. (2006). In S. Macedo, & Josiah Ober (Eds.), *Primates and philosophers: How morality evolved.* Princeton, NJ: Princeton University Press.
Dugatkin, L. A. (2002). Animal cooperation among unrelated individuals. *Naturwissenschaften*, *89*, 533—541.
Dugatkin, L. A. (2006). *The altruism equation: Seven scientists search for the origins of goodness.* Princeton, NJ: Princeton University Press.
Eccles, J. C. (1989). *Evolution of the brain: Creation of the self.* London: Routledge.
Egan, L. C., Santos, L. R., & Bloom, P. (2007). The origins of cognitive dissonance: Evidence from children and monkeys. *Psychological Science*, *18*, 978—983.
Elliott, C. (1996). *The rules of insanity: Moral responsibility and the mentally ill offender.* Albany: State University of New York Press.
Feschbach, S. (1964). The function of aggression and the regulation of aggressive drive. *Psychological Review*, *71*, 257—272.
Ford v. Wainwright, 477 U.S. 399. (1986).
Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, *68*, 5—20.
Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, *7*, 77—83.
Gardner, R., & Cory, G. A. (2002). *The evolutionary neuroethology of Paul MacLean: convergences and frontiers.* New York: Praege.
Gazzaniga, M. (2005). *The ethical brain.* New York: Dana Press.
Gazzaniga, M. (2007). My brain made me do it. In W. Glannon & Walter (Eds.), *Defining right and wrong in brain science: Essential readings in neuroethics. The Dana Foundation series on neuroethics* (pp. 183—194). Washington, DC: Dana Press.
Geen, R. G. (2001). *Human Aggression (Mapping Social Psychology) 2nd edition.* Berkshire: Open University Press.
Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, *2006*, 169—179.
Goodenough, O. R. (1998). Biology, behavior, and criminal law: Seeking a responsible approach to an inevitable interchange. *Vermont Law Review*, *22*, 263—294.
Goodenough, O. R. (2001). Mapping cortical areas associated with legal reasoning and moral intuition. *Jurimetrics*, *41*, 429—442.
Goodenough, O. R. (2004). Responsibility and punishment: Whose mind? A response. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, *359*, 1805—1809.
Greene, J., & Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, *359*, 1775—1785.
Greene, J., Nystrom, L., Engell, A., Darley, J., & Cohen, J. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*, 389—400.
Gruter, M. (1991). Behavior, evolution and the sense of justice. *American Behavioral Scientist*, *34*, 371—385.
Gruter, M. (1979). The origins of legal behavior. *Journal of Social and Biological Structures*, *2*, 43—51.
Hamilton, W. D. (1964). The genetical evolution of social behavior, I & II. *Journal of Theoretical Biology*, *7*, 1—52.
Harada, T., Itakura, S., Xu, F., Lee, K., Nakashita, S., Saito, D. N., et al. (2009). Neural correlates of the judgment of lying: A functional magnetic resonance imaging study. *Neuroscience Research*, *63*, 24—34.
Haushofer, J., & Fehr, E. (2008). You shouldn't have: Your brain on others' crimes. *Neuron*, *60*, 738—740.
Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*, 1362—1367.
Hinde, R. A. (2004). Law and the sources of morality. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, *359*, 1685—1695.
Hoffer, E. (1951). *The true believer: Thoughts on the nature of mass movements.* New York: Harper & Brothers.
Hoffman, M. B., & Goldsmith, T. H. (2004). The biological roots of punishment. *Ohio State Journal of Criminal Law*, *1*, 627—641.
Hsu, M., Anen, C., & Quartz, S. R. (2008). The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science*, *320*, 1092—1095.
Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty under decision making. *Science*, *310*, 1680—1683.
Illes, J. (2007). Empirical neuroethics. Can brain imaging visualize human thought? Why is neuroethics interested in such a possibility? *EMBO* (Supplement 1), S57—S60 Reports 8.
Johansson, P., Hall, L., Sikstrom, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, *310*, 116—119.
Jones, C. P. A. (2004). Law and morality in evolutionary competition (and why morality loses). *University of Florida Journal of Law and Public Policy*, *15*, 285.
Jones, O. D. (2000). On the nature of norms: Biology, morality, and the disruption of order.*Michigan Law Review*, *98*, 2072 Downloaded 12/08/08 from: http://ssrn.com/abstract=248391 or DOI: 10.2139/ssrn.248391
Jones, O. D. (2006). Law, evolution and the brain: Applications and open questions. In S. Zeki & O. R. Goodeneough (Eds.), *Law and the brain* (pp. 57—75). Oxford: Oxford University Press.
Kahane, G., & Savulescu, J. (2009). Brain damage and the moral significance of consciousness. *Journal of Medicine and Philosophy*, *34*, 6—26.
Kane, R. (1996). *The significance of free will.* Oxford: Oxford University Press.
Karli, P. (2006). The neurobiology of aggressive behaviour. *Comptes Rendus Biologies*, *329*, 460—464.
Katz, L. D. (2000). Toward good and evil: Evolutionary approaches to aspects of human morality. In L. D. Katz (Ed.), *Evolutionary origins of morality: Cross disciplinary perspectives* (pp. ix—xvi). Exeter, UK: Imprint Academic.
Keckler, C. N. W. (2006). Cross-Examining the brain: A legal analysis of neural imaging for credibility impeachment. *Hastings Law Journal*, *57*, 509—556.
Kingsbury, S. J., Lambert, M. T., & Hendrickse, W. (1997). A two-factor model of aggression. *Psychiatry*, *60*, 224—232.
Koenig, L. B., & Bouchard, T. J., Jr. (2006). Genetic and environmental influences on the traditional moral values triad—authoritarianism, conservatism, and religiousness—as assessed by quantitative behavior genetic methods. In P. McNamara (Ed.), *Where God and science meet: How brain and evolutionary studies alter our understanding of religionEvolution, genes, and the religious brain. Psychology, religion, and spirituality, Vol. 1* (pp. 47—76) Westport, CT, US: Praeger Publishers/Greenwood Publishing Group.
Konner, M. (2003). *The Tangled wing: Biological constraints on the human spirit.* New York: Holt Paperbacks.
Lerner, J. S., & Tiedens, L. Z. (2006). Portrait of the angry decision maker: How appraisal tendencies shape anger's influence on cognition. *Journal of Behavioral Decision Making*, *19*, 115—137.
Libet, B. (1993). *The neural time factor in conscious and unconscious mental events. Experimental and Theoretical Studies of Consciousness, Ciba Foundation Symposium #174.* Chichester: Wiley.
Libet, B. (1999). Do we have free will? In B. Libet, A. Freeman, & K. Sutherland (Eds.), *The volitional brain: Toward a neuroscience of free will.* Torverton, UK: Imprint Academic.
Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activities (readiness potential): The unconscious intention of a freely voluntary act. *Brain*, *106*, 623—642.
Little, T. D., Brauner, J., Jones, S. M., Nock, M. K., & Hawley, P. H. (2003). Rethinking aggression: A typological examination of the functions of aggression. *Merrill-Palmer Quarterly*, *49*, 343—369.
Loeber, R., & Hay, D. (1997). Key issues in the development of aggression and violence from childhood to early adulthood. *Annual Review of Psychology*, *48*, 371—410.
MacLean, P. D. (1990). *The triune brain in evolution: Role in paleocerebral functions.* New York: Plenum Press.
Mattson, M. P. (2003). *Neurobiology of aggression: Understanding and preventing violence.* Totowa, NJ: Humana Press.
McEllistrem, J. E. (2004). Affective and predatory violence: A bimodal classification system of human aggression and violence. *Aggression and Violent Behavior*, *10*, 1—30.
Mesterton-Gibbons, M., & Dugatkin, L. A. (1992). Cooperation among unrelated individuals: Evolutionary factors. *The Quarterly Review of Biology*, *67*, 267—281.
Mithen, S. (1996). *The Prehistory of the mind: The cognitive origins of art, religion, and science.* London: Thames and Hudson, Ltd.
Mobbs, D., Lau, H., Jones, O., & Frith, C. (2007). Law, responsibility, and the brain. *PLoS Biology*, *5*, 693—700.
Morris, S. G. (2007). Neuroscience and the free will conundrum. *The American Journal of Bioethics*, *7*, 20—22.
Morse, S. J. (2002). Uncontrollable urges and irrational people. *Virginia Law Review*, *88*, 1025—1064.
Morse, S. J. (2004a). New neuroscience, old problems: Legal implications of brain science. *Cerebrum*, *6*, 81—90.
Morse, S. J. (2004b). Reason, results and criminal responsibility. *University of Illinois Law Review*, *2*, 363—444.
Morse, S. J. (2006). Moral and legal responsibility in the new neuroscience. In J. Illes (Ed.), *Neuroethics: Defining the issues in theory, practice and policy* (pp. 29—36). Oxford: Oxford University Press.
Morse, S. J. (2007a). The non-problem of free will in forensic psychiatry and psychology. *Behavioral Sciences and the Law*, *25*, 203—220.

Morse, S. J. (2007b). Voluntary control of behavior and responsibility. *The American Journal of Bioethics*, *7*, 12—13.
Morse, S. J., & Hoffman, M. B. (2007). The uneasy entente between legal insanity and *mens rea*: Beyond Clark v. Arizona. *The Journal of Criminal Law and Criminology*, *97*, 1071—1150.
Moyer, K. E. (1976). *The psychobiology of aggression.* New York: Harper and Row Publishers.
Murphy, N. C., & Brown, W. S. (2007). *Did my neurons make me do it?: Philosophical and neurobiological perspectives on moral responsibility and free will.* Oxford: Oxford University Press.
Naqvia, N., Shiv, B., & Bechara, A. (2006). The role of emotion in decision making: A cognitive neuroscience perspective. *Current Directions in Psychological Science*, *15*, 260—264.
Nelson, R. J., & Trainor, B. C. (2007). Neural mechanisms of aggression. *Nature Reviews. Neuroscience*, *8*, 536—546.
Newberg, A., D'Aquill, E., & Rause, V. (2001). *Why religion won't go away: Brain science and the biology of belief.* New York: Ballantine Books.
O'Hara, E. A. (2004). How neuroscience might advance the law. *Philosophical Transactions of the Royal Society, Series B: Biological Sciences*, *359*, 1677—1684.
Olweus, D. (1979). Stability of aggressive reaction patterns in males: A review. *Psychological Bulletin*, *86*, 852—875.
Panchanathan, K., & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, *432*, 499—502.
Panetti v. Quarterman, 551 U.S. ___.(2007).
Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions.* New York: Oxford University Press.
Papez, J. (1937). A proposed mechanism of emotion. *Journal of Neuropsychiatry and Clinical Neurosciences*, *7*, 103—112.
Parrotta, D. J., & Giancola, P. R. (2007). Addressing "The criterion problem" in the assessment of aggressive behavior: Development of a new taxonomic system. *Aggression and Violent Behavior*, *12*, 280—299.
Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2004). Grasping the intentions of others: The perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience*, *16*, 1706—1716.
Penry v. Lynaugh, 492 U.S. 302. (1989).
Pereboom, D. (2001). *Living without free will.* Cambridge: Cambridge University Press.
Pettit, G. S. (1997). The developmental course of violence and aggression: Mechanism of family and peer influence. *The Psychiatric Clinics of North America*, *20*, 283—299.
Phillips, J. K. G., & Woodman, R. E. (2007). *The insanity of the mens rea model: Due process and the abolition of the insanity defense (September 4, 2007)* Downloaded 12/18/08 from SSRN: http://ssrn.com/abstract=1012104
Prehn, K., Wartenburger, I., Meriau, K., Scheibe, C., Goodenough, O. R., Villringer, A., et al. (2008). Individual differences in moral judgment competence influence neural correlates of socio-normative judgments. *Social Cognitive and Affective Neuroscience*, *3*, 33—46.
Premack, D. G., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*, 515—526.
Prinz, J. J. (2008). Is morality innate? In W. Sinnott-Armstrong (Ed.), *Moral psychologyThe evolution of morality: Adaptations and innateness*, *Vol. 1* (pp. 367—406). Cambridge, MA: MIT Press.
Rakos, R. R., Laurene, K. R., Skala, S., & Slane, S. (2008). Belief in free will: Measurement and conceptualization innovations. *Behavior and Social Issues*, *17*, 20—39.
Raymond, P. E. (1936). The origin and rise of moral liability in Anglo-Saxon criminal law. *Oregon Law Review*, *15*, 93—117.
Reep, R. L., Finlay, B. L., & Darlington, R. B. (2007). The Limbic system in mammalian brain evolution. *Brain Behavior and Evolution*, *70*, 57—70.
Ridely, M. (1996). *The origins of virtue.* New York: Viking.
Rizzolatti, G., & Craighero, L. (2004). The mirror–neuron system. *Annual Review of Neuroscience*, *27*, 169—192.
Robertson, D., Snarey, J., Ousley, O., Harenski, K., Bowman, F. D., Gilkey, R., & Kilts, C. (2007). The neural processing of moral sensitivity to issues of justice and care. *Neuropsychologia*, *45*, 755—766.
Rockenbach, B., & Milinski, M. (2006). The efficient interaction of indirect reciprocity and costly punishment. *Nature*, *444*, 718—723.
Roper v. Simmons, (03–633) 543 U.S. 551.(2005).
Roskies, A. L. (2008). Response to Sie and Wouters: A neuroscientific challenge to free will and responsibility? *Trends in Cognitive Sciences*, *12*, 4.
Sayre, F. B. (1932). *Mens rea.* (Citing De legibus et consuetudinibus angliae). *Harvard Law Review*, *45*, 974—1026.
Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in 'theory of mind'. *NeuroImage*, *19*, 1835—1842.
Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science*, *17*, 692—699.
Scheibel, A. B., & Schopf, J. W. (1997). *The origin and evolution of intelligence.* Sudbury, MA: Jones and Bartlett Publishers.
Seymour, B., Singer, T., & Dolan, R. (2007). The neurobiology of punishment. *Nature Reviews. Neuroscience*, *8*, 300—311.
Sie, M., & Wouters, A. (2008). The real challenge to free will and responsibility. *Trends in Cognitive Sciences*, *12*, 3—4.
Siegel, A. (2004). *Neurobiology of Aggression and Rage.* New York: Informa Healthcare.
Simon, A. (1997). *Models of bounded rationalityEmpirically grounded economic reason*, *Vol. 3*. Cambridge, MA: MIT Press.
Singer, T. (2007). The neuronal basis of empathy and fairness. *Novartis Foundation Symposium*, *278*, 20—30.
Slobogin, C. (2005). The civilization of the criminal law. *Vanderbilt Law Review*, *58*, 121—168.
Slote, M. (1990). Ethics without free will. *Social Theory and Practice*, *16*, 369—383.
Spence, S. A., Hunter, M. D., Farrow, T. F. D., Green, R. D., Leung, D. H., Hughes, C. J., et al. (2006). A cognitive neurobiological account of deception: Evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society, Series B*, *39*, 1755—1762.
Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior.* Cambridge, MA: Harvard University Press.
Stake, J. E. (2001). Can evolutionary science contribute to discussions of law? *Jurimetrics*, *41*, 379—384.
Stanford v. Kentucky, 492 U.S. 361 (1989).
Stefánsson, H. (2007). The biology of behaviour: Scientific and ethical implications. *EMBO*, S1—S2 reports 8, Suppl. 1.
Striedter, G. F. (2005). *Principles of brain evolution.* Sunderland: Sinauer Associates.
Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes of forthcoming perception in the frontal cortex. *Science*, *314*, 1311—1314.
Swanson, K.A. (2002). Criminal law: *Mens rea* alive and well: Limiting public welfare offenses. In re. C.R.M. William Mitchell Law Review, 28, 1265–1282.
Takahashi, H., Kato, M., Matsuura, M., Koeda, M., Yahata, N., Suhara, T., & Okubo, Y. (2008). Neural correlates of human virtue judgment. *Cerebral Cortex*, *18*, 1886—1891.
Tancredi, L. (2005). *Hardwired behavior: What neuroscience reveals about morality.* Cambridge, UK: Cambridge University Press.
Taylor, C., & Dennet, D. (2001). Who's afraid of determinism? Rethinking causes and possibilities. In R. Kane (Ed.), *Oxford handbook of free will* (pp. 257—277). New York: Oxford University Press.
Terry, D. A. (2002). Don't forget about reciprocal altruism: Critical review of the evolutionary jurisprudence movement. *Connecticut Law Review*, *34*(477), 485—487.
Thompson v. Oklahoma, 487 U.S. 815 (1988).
Torr, J. D., & Egendorf, L. K. (Eds.). (2000). *Problems of death: Opposing viewpoints.* San Diego, CA: Greenhaven Press.
Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*, 35—57.
van Hooff, J. C. (2008). Neuroimaging techniques for memory detection: Scientific, ethical, and legal issues. *The American Journal of Bioethics*, *8*, 25—26.
Victoroff, J. (2009). Human aggression. In B. J. Sadock & V. A. Sadock (Eds.), *Kaplan and Sadock's comprehensive textbook of psychiatry*, IXth Ed. Philadelphia: Lippincott Williams and Wilkins.
Vincent, N. (2008). Responsibility, dysfunction and capacity. *Neuroethics*, *1*, 199—204.
Waldbauer, J. R., & Gazzaniga, M. S. (2001). The divergence of neuroscience and law. *Jurimetrics*, *41*, 357—364.
Walsh, A. (2000). Evolutionary psychology and the origins of justice. *Justice Quarterly*, *17*, 841—864.
Wegner, D. (2002). *The illusion of conscious will.* Cambridge, MA: MIT Press.
Wigley, S. (2007). Automaticity, consciousness and moral responsibility. *Philosophical Psychology*, *20*, 209—225.
Wilson, J. Q. (1993). *The Moral Sense.* New York: Free Press.
Wolf, S. M. (2008). Neurolaw: The big question. *The American Journal of Bioethics*, *8*, 21—22.
Wright, R. (1994). *The moral animal: The new science of evolutionary psychology.* New York: Pantheon.
Young, L., & Saxe, R. (2008a). The neural basis of belief encoding and integration in moral judgment. *Neuroimage*, *40*, 1912—1920.
Young, L., & Saxe, R. (2008b). An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience*, *18*, 803—817.
Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E. D., et al. (2009). The neural basis of human social values: Evidence from functional MRI. *Cerebral Cortex*, *19*, 276—283.