Topics in the Neurobiology of Aggression: Implications to Deterrence

February 2013

Contributing Authors: Robert M. Sapolsky, Ph.D., Stanford University, John A. Gunn, Ph.D., Stanford University, Cynthia Fry Gunn Ph.D., Stanford University, Allan Siegel, Ph.D., New Jersey Medical School, Jordan Grafman, Ph.D., Rehabilitation Institute of Chicago, Pamela Blake, M.D., Memorial Hermann Northwest Hospital, Peter K. Hatemi, Ph.D., Pennsylvania State University, Rose McDermott, Ph.D., Brown University, Anthony C. Lopez, Ph.D., Washington State University, Paul J. Zak, Ph.D., Claremont Graduate University, James Giordano Ph.D., MPhil., Georgetown University Medical Center, Roland Benedikter Ph.D., DrPhil., Stanford University, & Lieutenant General Robert E. Schmidle, Jr., United States Marine Corps

Editors: Dr. Diane DiEuliis (HHS) and Dr. Hriar Cabayan (DOD)

A Strategic Multi-Layer (SMA) Periodic Publication

Executive Summary

This paper provides a series of selected topics on the neurobiological basis of aggression, in the interest of introducing this field of science to the community of experts in deterrence. Thus the topics covered are those of most relevance to inform a deterrence community of how neurobiological considerations may be incorporated as potentially useful tools in deterrence practice. The material is presented in 3 Parts: A summary of the basic neurobiology, a few chapters on behavioral impacts of that neurobiology, and finally a "systems summary" that integrates these issues for readers. Topics include the role of genes and environment, group vs. individual aggressive behaviors, the role of trust and altruism, reward/punishment, and premeditated vs. impulsive aggression, among others. The paper also discusses the continuum of scientific evidence base, from the basic neurobiology of reactive aggression, to pre-meditated aggression and what connections can be understood and conjoined to the psychological and behavioral evidence base.

Key insights provided by contributing authors to this white volume that are of particular relevance to the operational community include:

- It is not possible to understand the biology of behavior without understanding the context in which that biology occurs, as well as the society in which that individual dwells. This is true in our understanding of aggression; there is no highly accurate means of identifying individuals likely to commit an impulsive or planned violent act. *The context in which aggression and violence occur can be modified much more easily than identifying individuals likely to commit aggressive act; by manipulating context, society may reduce aggression by individuals indirectly.*

- Much aggression is motivated by conflict between in-groups and out-groups. An understanding of genetic and environmental factors can elucidate pathways toward aggression and begin to explain how various environmental factors such as media, propaganda, or informal mechanisms of narrative messaging, can be used to manipulate the neurobiological mechanisms that inform the psychological architecture of susceptible individuals. *In that context, foreign policies which overtly impose governance or values alien to local cultures, may constitute provocations to violence.*

- Within groups, punishment and reward cannot be understood outside the context of cooperation. Cooperation is stable when defectors can be identified, excluded and/or punished, and when prospective cooperators can be identified, engaged, and rewarded through cooperative exchange. *Research indicates reward may function less effectively as a behavior-changing strategy, but may function more effectively as a behavior-sustaining strategy.*

- Punishment in the context of group conflict cannot be understood absent the evolutionary logic of warfare between groups in an ancestral environment that was "offense dominant". The "secure retaliatory force" that nuclear strategists argue is necessary for equilibrium in the nuclear age is nothing but a euphemism for "guaranteed vengeance," in which states promise a punishment that is greater than the benefits of striking first.

- States where the rule of law is weak can beget societies characterized by "culture of honor" traditions, in which, in the *absence of capable and legitimate third-party enforcement, reputation for disproportionate retaliation/punishment becomes the most effective safeguard against personal violence.*

- *Deterrence as a concept may be a long-learned part of our psychology*. Because challenges, predators, or out-group threat have faced humans for millennia, analyzing the notion of deterrence from the perspective of evolutionary models may prove helpful. Rational actors have an interest in settling things with threats but without the use of violence. Vengeance is certain to be provoked by an attack, deterrence kicks in when the initiators cannot be absolutely sure that they'll be successful.

- *Neuroscientific studies of human behavior suggest aggression is a useful but costly strategy and people have a tendency to cooperate in many situations.* This has been shown to be chemically regulated, and cooperative behaviors are just as "natural" as aggressive ones. The "Golden Rule" exists in every culture and reveals our essential social nature; this is naturally threatened in situations of threat or high stress.

- Neuroscience and technology offer a viable and potential value of in programs of national security, intelligence and defense. However this will necessitate an acknowledgement of the actual capabilities and limitations of the neuroS/T used – and importantly the ethico-legal issues generated by apt or inapt use, or blatant abuse.

- Context is critically important to our understanding of human behavior and biology is only part of what makes up our human selves and defines us as persons. An integrated approach, one that takes into account the influence of the empirical sciences as well as a social psychological framework, gives us the most holistic understanding of human behavior and produce the greatest improvement in our understanding of terrorism.

Topic Overview

In his introduction, Robert Sapolsky defines "aggression" in the fullest sense as harming, attempts to harming, and thoughts about harming. Two brain regions dominate this. The first is the amygdala, an ancient structure in the "limbic system" (the "emotional" part of the brain).

The second is the frontal cortex (FC). It is more complex in humans than in other species, is the most recently evolved part of our brain, and is the last to fully mature (remarkably, in the mid-20s). There are tremendous individual differences in FC function among people; many of differences arise from differing early life experiences. The limbic system/FC contrast can, in a thoroughly simplistic way, be thought of as a contrast between emotion and thought. A key issue is that genes have far less to do with aggression than is often assumed; genes are not about inevitability, rather, they are about proclivities. In conclusion, Professor Sapolsky states that amid this emphasis on biology, ultimately, it is not possible to understand the biology of behavior without understanding the circumstances of the individual in which that biology occurs, as well as the society in which that individual dwells.

In Part 1 (Aggression and the Brain), Allen Siegel discusses the two forms of aggressive behavior: predatory or premeditated aggression and affective or impulsive aggression. He points out that it is often difficult to detect any response patterns that could be used to predict the onset of predatory aggression – the behavior occurs with no perceived threat, is purposeful and planned, and appears to be devoid of conscious awareness of emotion.

Jordan Grafman and Pamela Blake provide further insights into the aggressive brain. They focus on laboratory work examining the key brain regions involved in impulsive aggressive behavior and control, the types of environmental exposures that could influence aggressive behavior, and the interaction of genetic predisposition, brain damage, and aggressive behavior. They report that specific brain lesions can affect impulsive aggressive behavior, and genetic predisposition can affect aggression as long as the key brain area mediating that effect is not damaged. There is no guarantee of accurately identifying people likely to commit an impulsive or planned violent act, rather, the context in which aggression and violence occur can be modified much more easily than identifying individuals likely to commit aggressive acts. By manipulating context, society can reduce aggression by individuals indirectly.

Peter Hatemi and Rose McDermott point out that much aggression is motivated by conflict between in-groups and out-groups. Aligned with other authors in this white paper, they note that studies of genetic and environmental characteristics can elucidate pathways of aggression. They can provide insights into how various environmental factors, such as the media, propaganda, and informal mechanisms of narrative messaging, can be used by ourselves, allies and adversaries to manipulate the neurobiological mechanisms that inform the psychological architecture of susceptible individuals. From this perspective, US interests may be better protected through the development of strategies to manipulate environmental triggers, and creation of interventions to address the human psychological architecture that responds to threat with aggression. Lastly, provocation is important in potentiating violence. Foreign policies which try to impose governments, institutions, or values alien to local cultures, are likely to be understood as constituting such provocations.

In Part II, some implications of aggressive behavior are presented. Anthony Lopez discusses the human psychology of reward and punishment in light of evolutionary pressures, and in the context of both in-group and out-group behavior. Within groups, punishment and reward

cannot be understood outside the context of cooperation. Cooperation is stable when defectors can be identified, excluded and/or punished, and when prospective cooperators can be identified, engaged, and rewarded through cooperative exchange. Empirical data indicates that when punishment is available, cooperation is often stable and free-riding is deterred. Although research on the role of reward in promoting cooperation is relatively lacking, there are indications that reward may function less effectively as a behavior-*changing* strategy, but may function more effectively as a behavior-*sustaining* strategy. Punishment between individuals and groups takes the form of a withdrawal of benefits or the conferral of costs. Punishment in the context of group conflict cannot be understood absent the evolutionary logic of group warfare. Evolutionarily, warfare has most often taken the form of lethal raiding and it has occurred in an ancestral environment that was "offense dominant". Lopez notes that the "secure retaliatory force" (touted by nuclear strategists as necessary for equilibrium in the nuclear age) is actually a euphemism for "guaranteed vengeance," in which states promise a punishment that is greater than the benefits of striking first. States with weak rule of law beget societies characterized by "culture of honor" traditions, in which, in the absence of capable and legitimate third-party enforcement, reputation for disproportionate retaliation/punishment becomes the most effective safeguard against personal violence.

Rose McDermott and Peter Hatemi provide further insights into the role of emotion in decision making around violence. By exploring the foundations of human psychological and neurobiologically informed notions of threat and deterrence, we can begin to leverage our own biology in service of our very survival through recognition of those environmental cues and triggers which both instigate and extinguish our desires for aggression and cooperation. They discuss two key topics relevant to the theory of deterrence; namely the "Psychology of first strike, Coalitionary Humans, and Maximum Response" and "Leadership and Group Dynamics". Following similar arguments advanced by Anthony Lopez, they argue that before nuclear weapons appeared, deterrence as a concept was naturally built into our psychology. Because humans have faced challenges, predation, and out-group threats for millennia, analyzing the notion of deterrence from the perspective of evolutionary models may prove useful; examining the genetic and biological mechanisms which precipitate our recognition and response to threat can inform our understanding of how to create more accurate signals and more effective responses. Rational actors may negotiate by using threats, without the use of violence. Where vengeance is certain to be provoked by an attack, deterrence kicks in when the initiators cannot be absolutely sure that they'll be successful. The section on "Leadership and Group Dynamics" provides reasoning that evolutionary and neurobiological perspectives have served to enlighten aspects of leadership beyond that of traditional models. They propose that a more neurobiologically informed understanding of individual dispositions and personal psychology may reveal triggers that cue a particular individual to respond in a hostile as opposed to conciliatory manner in the face of threat.

In the last chapter in Part II, Paul Zak focuses on the role of hormones on the reduction of aggression. Episodes of aggression, especially repeated aggression by the same individual, are due to combinations of, and interactions between, genes, brains, history, and environments. Neuroscientific studies of human behavior suggest aggression is a useful but costly strategy and

most people have a strong bias to cooperate in many situations. Zak demonstrates this is affected by the neuroactive hormone oxytocin (OT). OT appears to function as a chemical regulator that mediates prosocial behaviors by signaling that another person is safe or familiar, even if the other person is a stranger. One can make the case that cooperative behaviors are just as "natural" as aggressive ones, that cooperation with strangers is a typical human behavior, and that conflict among strangers may not be the norm. He reports that laboratory "trust games" capture, in an objective way, the notion of the Golden Rule: if you are nice to me, I'll be nice to you. Of hundreds of people tested in a variety of cultures, roughly 95% of individuals reciprocate trust. The Golden Rule exists in every culture on the planet and reveals our essential social nature. It appears that OT is largely responsible for reciprocation by sending a safety signal motivating nice with nice. He goes on to point out the role of the environment and states that high stress can inhibit OT and put us into solitary or socially narrow survival mode. Environments that are unsafe, new, competitive, aggressive, or unpredictable induce greater release of certain hormones and thereby inhibit prosocial behaviors, especially such behaviors towards strangers.

In Part III, James Giordano shifts focus towards weaving the basics of neuroscience and technology ("neuroS/T") into social and cognitive aspects, creating a systems understanding. In a paper entitled "Toward a Systems Continuum: On the Use of Neuroscience and Neurotechnology to Assess and Affect Aggression, Cognition and Behavior" he argues that neuroscience has assumed a prominent role in shaping views of the human being, human condition, and human relationships. Neuroscientific discoveries continue to challenge and promote a re-examination of socially-defined ontologies, values, conventions, norms and mores, and the ethico-legal notions of individual and social good. As a potential analogous framework for "neurodeterrence" he introduces the concept of neuroecology - the study individuals' neural systems, embedded in groups and environment(s) framed by time, place, culture and circumstance. Defining the neural bases of such biological-environmental interactions may yield important information about factors that dispose and foster various actions - including cooperation, conflict, aggression and violence. This emphasizes the viability and potential value of neuroS/T in programs of national security, intelligence and defense (NSID). He reiterates that any consideration of the possible use of neuroS/T for NSID would require acknowledgement of the actual capabilities and limitations of the neuroS/T used – and importantly, the ethico-legal issues generated by apt or inapt use, or blatant abuse. Otherwise there is a real risk that neuroscientific outcomes and information may be misperceived, and misused to wage arguments that are inappropriate or fallacious. He questions whether ethico-legal systems are in place and realistic and mature enough to guide, direct and govern such possible use and/or non-use. He goes on to state there is the need to develop stringent technical and ethico-legal guidelines and standards for such use of neuroS/T. He advocates a dedication to both ongoing neuroS/T research, and full content ethico-legal address, analyses and articulation of the ways that these approaches may be used, misused and/or abused in contexts of national security, intelligence and defense.

Neuroscience and technology offer a viable and potential value of in programs of national security, intelligence and defense. However this will necessitate an acknowledgement of the

actual capabilities and limitations of the neuroS/T used – and importantly the ethico-legal issues generated by apt or inapt use, or blatant abuse.

LtGen Robert Schmidle (USMC) in his closing editorial comment chapter entitled “An Integrated Approach to Understanding Human Behavior” proposes some final ideas in line with other contributors to this white volume. He stresses that context is critically important to our understanding of human behavior, and biology is only part of what makes up our human selves and defines us as persons living in a given society and culture. He discusses the terrorist as a behavioral example of purposeful aggression and violence. He advocates an integrated approach, one that takes into account the influence of the empirical sciences as well as a social psychological framework, gives us the most holistic understanding of human behavior. He stresses the need to develop a conceptual framework within which to conduct empirical investigations that includes the relevant cultural and historical context. In this instance, both kinds of knowledge, scientific, (i.e. empirical) and philosophic, (i.e. conceptual) are necessary for our understanding of human actions. He concludes by stating that while we may not ever definitively answer the question of responsibility for the development of a terrorist, for example, an integrated approach that takes into account both biology and psychology will produce the greatest improvement in our understanding of terrorism.

Introduction to the Neurobiology of Aggression

Robert M. Sapolsky
John A. and Cynthia Fry Gunn Professor
Department of Biology, Stanford University
Departments of Neurology and Neurosurgery, Stanford University Medical School
and
Institute of Primate Research
National Museums of Kenya

If you are a neuroscientist, a central premise is that it is not possible to understand behavior, include human behavior, and even abnormal human behavior, without biology. But at the same time, another central premise must be that you're not going to understand behavior if you think that biology will explain everything.

In this overview, I summarize the biology of aggression, and how it can intersect with psychological and social. Given its brevity, this is obviously going to be hugely simplifying.

The issue of definitions

It is obligatory to start with the challenge of definitions. This is certainly the case with studying humans, where you must distinguish between violence and aggression as well as between actions and internal psychological states, and where the behaviors can range from hand-to-hand combat to directing a drone to kill someone whose face is never seen. Challenges occur as well when studying animals, where you cannot readily access its internal psychological states, and where aggression is easily confused with predatory behavior. To sidestep these subtleties, I will use "aggression" in the fullest sense – harming, attempts to harming, and thoughts about harming.

Neurobiological aspects

Two brain regions dominate this subject. The first is the amygdala, an ancient structure in the "limbic system" (the "emotional" part of the brain). Supporting evidence for this conclusion includes: a) neurons in the amygdala become active when aggression occurs; b) if the region is damaged, levels of aggression plummet; c) if the amygdala is electrically stimulated, aggression occurs; d) rare tumors in the amygdala produce aggression. These findings are based on a huge number of studies of humans and other species. The outflow from the amygdala (i.e., the brain regions that it projects to directly or indirectly) includes regions that initiate the actual behaviors as well as support the behavior by doing things like increasing heart rate and blood pressure.

The amygdala is one of the brain regions most sensitive to the actions of testosterone, as well as of a class of stress hormones. More on this later.

Conceptually, a critical point about the amygdala's role in aggression is that it is also the brain structure most central to fear and anxiety. This conclusion is also based on an enormous literature. I think it is fair to say that in a world in which no organism need be afraid, the amygdala will be activating aggression-relevant circuitry at a much lower rate.

The other key region of the brain is the frontal cortex (FC). It is more complex in humans than in other species, is the most recently evolved part of our brain, and is the last to fully mature (remarkably, in the mid-20s). The FC is key to executive decision-making, long-term planning, gratification postponement, impulse control and emotional regulation. Basically, the FC makes you do the harder thing when it's the right thing to do. In the cognitive realm, it facilitates focus amid distractions, and inhibits "overlearned responses" (for example suppressing the urge to recite the months in sequence when trying to rapidly recite them backwards). And it is key to regulating social behaviors, such that damage to the FC can disinhibit inappropriate behaviors; as an example, violent sociopaths have decreased metabolic rates and abnormal function in the FC.

Much of the FC's regulation of emotional behavior arises from its ability to inhibit the amygdala. The amygdala, in turn, sends inhibitory projections to the FC; this is a means by which, metaphorically, a violent passion can overcome reason. Commensurate with this, there is an inverse relationship between activity levels in the amygdala and FC.

Importantly, there are tremendous individual differences in FC function among people; many of differences arise from differing early life experiences; more on this later.

Finally, the limbic system/FC contrast can, in a thoroughly simplistic way, be thought of as a contrast between emotion and thought. Increasing evidence suggests that in circumstances of rapid decision-making in an aroused context, the limbic system first "decides" what action will be taken, followed by cortical regions rationalizing that decision.

Sensory regulation of aggression-related neurobiology

Obviously, regions like the amygdala and FC receive information about the outside world; how else can your brain tell your heart to race if you must sprint from a lion?

Two important points in this domain:

- These structures can be responsive to subliminal stimuli (i.e., stimuli that are so rapid or faint that there is no conscious awareness of them).
- A stressed, aroused state causes a critical change in sensory inputs to the amygdala.

Normally, sensory information is first processed in the sensory cortex. For example, the visual cortex turns pixels into dots and then lines, and then three-dimensional precepts, until a coherent image is passed on to the associative cortex. If the visual information is of emotional relevance, signals are then sent to the amygdala. By neurobiological standards, this is a glacial process.

During periods of high arousal and stress, a second pathway of sensory inputs is also utilized. In effect, a shortcut occurs before sensory information reaches the cortex; this branch sends information directly to the amygdala. Thus, the amygdala can gain emotionally relevant information before the cortex processes it. Critically, by bypassing the cortical processing, the information sent to the amygdala is often inaccurate. This gives rise to the scenario of someone shooting before realizing that the handgun the other person is holding is actually a cell phone.

Hormonal effects on pertinent parts of the brain

All hormones can alter brain function. I will touch on only two.

One is a class of hormones called glucocorticoids that are secreted during stress. To summarize briefly, glucocorticoids a) help activate that short cut sensory pathway to the amygdala; b) increase the excitability of the amygdala; c) decrease the excitability of the FC. This is a mechanism by which judgment can be impaired by stress.

The other hormone is, of course, testosterone; if you remember only two things from this document, one is that testosterone's role in aggression has been vastly exaggerated. To summarize, in its normal range, testosterone does not cause aggression but, rather, lowers the threshold for social stimuli to trigger aggression. Some examples:

- in the amygdala, testosterone does not excite neurons. Instead, it causes neurons to be more excited if and only if they are being stimulated by some other input.
- if you look at faces with expressions ranging from friendly to menacing, testosterone makes you more likely to interpret emotionally ambiguous ones as menacing.
- if testosterone is administered to a middle-ranking male monkey, it increases his level of aggression. However, this takes the form of him being more aggressive towards hierarchical subordinates, rather than now challenging males who dominate him. In other words, testosterone does not alter the hierarchy but, instead, exaggerates it.
- castration drastically decreases aggression in all species examined. But critically, aggression does not disappear entirely. Instead, the more experience the individual had prior to castration, the less it declines afterward; in other words, the residual aggression has little to do with hormones and much to do with social factors.

Early experiences and the pertinent neurobiology

This domain can be summarized as follows: social science has demonstrated that adverse childhood events, the likes of trauma, abuse and loss, can drastically influence adult aggressive and empathic behavior; enough neuroscience is known by now to reveal the mechanisms by which this occurs, or at least to construct plausible models. As one example, by the age of kindergarten, the stress of low socioeconomic status produces a less developed FC in a child, a predictor for poor impulse control and executive function many years later. Three points from this large field:

- Environment does not begin at birth. Prenatal environment (for example, levels of stress hormones in the mother's circulation) influences the fetus' brain development.
- It is understood on a molecular level, how early experience can cause life-long changes in the brain (a trendy subject called "epigenetics").
- Many of those epigenetic changes can be lessened in the right adult setting.

Genes and aggression

If you remember only two things from this document, the second is that genes have far less to do with aggression than is often assumed. Genes are not about inevitability; they are about proclivities. Some key points:

- The field of "behavior genetics," which heavily depends on studies of twins or adopted individuals, has generated some very high measures of heritability of aggression. However, the field is fraught with major methodological and conceptual problems.
- Genes have different effects in different environments. An example having a particular version of a brain enzyme called Monoamine Oxidase (MAO-b) increases the incidence of anti-social behavior in young adults – but only if accompanied by a history of childhood abuse. Another example concerns a celebrated study about a family with many members with an MAO-b mutation and a history of violence. However, among other relatives, the mutation was associated with different abnormalities (e.g., kleptomania, or exposing themselves).
- Many genes linked to aggression in animal studies turn out to influence something very different (for example, lowering pain thresholds, so that animals are more likely to lash out aggressively in response to a painful stimulus, and that increase emotional reactivity in general, rather than solely in the realm of aggression).

The evolution of aggression

Organisms do not behave for the good of the species; they behave to maximize the number of copies of their genes in the next generation. This has readily given rise to the false idea that

evolution is solely about selection for pure self-interest. This is best summarized with the famous sound bite that evolution selects for "Selfish Genes."

In reality, the picture is far more complicated in three broad ways:

- "Individual selection" (i.e., selection for individuals to act in pure self-interest) can certainly selection for aggression. However, there are a variety of alternative strategies available. As one example, among primate species with intense hierarchical male-male competition for access to females, a consistent subset of individuals opt out of the competition. Instead, they develop stable affiliative relationships with females (something that, appropriately, has been termed "friendships"); these friendships are not necessarily Platonic, but involve furtive matings. Paternity testing shows that this strategy is a viable alternative for passing on copies of genes.
- "Kin selection" is built around the fact that relatives share genes, with more shared among closer relatives. This produces an iconic fact – from the standpoint of passing on copies of genes, it is equivalent to reproduce once or to make it possible for a full sibling (who shares 50% of genes) to reproduce twice. This is summarized in the aphorism – "I'll lay down my life for two brothers or eight cousins." From this comes the evolutionary drive towards cooperation among relatives in numerous species. One consequence of this is shown in a study of baboons, where the more female relatives a female has in her troop, the better the chances were of her infant surviving.
- Among numerous species, cooperation occurs among unrelated individuals (for example, among vampire bats, unrelated females feed each other's babies). A key question in evolutionarily biology is how such cooperation ever emerges among a sea of non-cooperating individuals. This is the domain of a rich literature concerning game theory, the mathematical analysis of when the optimal strategy is to cooperate and when not. Much of this work is focused on issues of reciprocity. Out of this has come insights into circumstances that favor the emergence of cooperation, as well as of detection and punishment of non-cooperators.

Thus, evolutionary self-interest occurs amid circumstances that select against aggression, and natural selection can reward cooperation among relatives and even unrelated individuals.

Conclusions

This simplified overview generates a number of broad conclusions:

- Biological factors often do not cause aggression as much as modulate responses to other stimuli.
- Be skeptical of biological explanations of behavior that are predominately about a single factor (especially genes or testosterone). All of these levels of analyses are intertwined.
- The brain is rarely purely autonomous, "making up its own mind" on its own. Instead, its function is tightly regulated by the outside world.

- Almost always, a hormone such as testosterone does not have an effect. Instead, it has a particular effect in a particular environment. Moreover, environment strongly regulates hormone secretion.
- Similarly, a gene typically does not have an intrinsic effect by itself, but instead, has an effect specific to specific environments. And similarly, environment is what most strongly regulates gene activity.
- Evolution is about maximizing the number of copies of genes passed on to the next generation, and this often involves selection against aggression and for cooperation.

Amid this emphasis on biology, ultimately, it is not possible to understand the biology of behavior without understanding the circumstances of the individual in which that biology occurs, as well as the society in which that individual dwells.

Part I: Aggression and the Brain

Chapter 1: Premeditated vs. impulsive aggression

Allan Siegel, Ph.D.
New Jersey Medical School
Siegel@umdnj.edu

Aggressive behavior may be defined as a type of behavior that threatens harm or leads to or causes harm, destruction or damage to another organism. In this context, aggression is not a unitary phenomenon, but instead, reflects a variety of different behavioral processes that are contained under a single heading. A variety of animal research models of aggressive behavior have been utilized, which include fear-induced, maternal, inter-male, irritable, sex-related, territorial, resident-intruder, affective defense and predatory aggression. With the exception of predatory aggression, these models of aggressive behavior share the following common features: they reflect a perceived or real threat, are aversive to the organism, are impulsive, display sympathetic signs, and are defensive in nature. Thus, these models can be reduced to the general category of affective (or defensive) aggression. In contrast, predatory aggression is quite different from the others, in particular because it requires planning, shows few autonomic signs, and is positively reinforcing to the organism. Thus, this chapter addresses the nature of two forms of aggressive behavior: predatory or premeditated aggression and affective or impulsive aggression. The first phase of the present discussion summarizes the behavioral and neurobiological characteristics of aggression that correspond to those linked to feline predatory aggression or affective defense, while the second phase briefly summarizes corresponding forms of aggression in humans..

Overview of affective (impulsive) and predatory (premeditated) aggression in the cat

Affective (impulsive) aggression occurs in nature in response to the presence of, or perceived presence of, a threatening stimulus such as another species within its territory. It is characterized by arching of the back, retraction of the ears, piloerection, pupillary dilatation, marked hissing, and striking of the target species with the forepaw at the threatening object. This form of aggressive behavior can also be elicited by electrical or chemical stimulation of the midbrain periaqueductal gray (PAG) and medial hypothalamus (Siegel, 2005; Siegel et al., 1999, 2007; Siegel and Victoroff, 2009).
The behavioral properties of feline affective defense can be listed as follows:

- intense sympathetic arousal
- vocalization, arching of back, marked pupillary dilatation, retraction of ears, piloerection
- impulsive (immediate) reaction to a perceived threat
- includes most categories of animal aggression

Predatory (quiet biting) attack is quite different from other forms of aggression. In the animal kingdom, it is manifest as hunting behavior. The attack is planned and is preceded by stalking of specific prey object of another species followed by biting of the back of the neck until it kills the prey. In contrast to defensive rage, the cat displays few autonomic signs aside from mild pupillary dilatation. Predatory attack is induced following electrical stimulation of the lateral hypothalamus or ventrolateral aspect of the PAG and ceases immediately following termination of stimulation.

The behavioral properties of predatory (premeditated) aggression can be summarized as follows:

- little sympathetic arousal
- triggered by presence of a prey object
- planned, clearly directed form of attack

Corresponding forms of aggression in humans

It is instructive to attempt to compare affective defense in animals with what some authors refer to as equivalent response forms in humans [Meloy, 1988; Vitiello et al., 1990]. These authors characterize affective defense behavior as impulsive, destructive to the object of the aggressor, and typically aversive to the aggressor. In addition, they further indicate that this response pattern included poor modulation of behavior and high autonomic arousal. Questionnaire test items closely related to affective defense include the following: aggression that was unplanned, an individual who was totally out of control during the aggressive act, aggression that had no purpose, the person is exposed to physical harm during the time when he is aggressive, and an individual who damages his own property during the act of aggression [Vitiello et al., 1990].

The characteristics associated with affective defense behavior in humans have been independently described in detail by Meloy [1988]. Several of these characteristics clearly overlap with those described by Vitiello. These include: an intense sympathetic arousal, an affective attack based upon a real or perceived (which could be delusional) threat to the person, and an immediate (i.e., impulsive) response to the threat stimulus. Meloy extends his analysis by adding other properties of this form of aggression. Specifically, the goal object of affective defense is to reduce or eliminate the threat object from the environment, and thus presumably reduce the level of tension. A second feature is that the person or animal displaying affective aggression can also easily show displacement of the perceived threat from one object to another. For example, during the time period when affective defense is expressed, the aggressor may easily attack a third person who accidentally entered the room instead of his original target. A third characteristic is that, as a result of sympathetic arousal, the behavioral repertoire is typically limited in time to events of short duration (usually lasting no longer than a few seconds or a minute). A fourth characteristic is that the aggressor elicits a ritualized or stereotyped posture displaying defense and attack prior to the initiation of the actual attack. Such posturing could take the form of a clenching of the fists, other gestures and

the use of obscene language. Ostensibly, the goal of such behavior is to demean the presence of the existing threat. Finally, the individual is able to subjectively experience the emotional state such as fear or anger that occurs during the time at which affective defense is elicited. In animals as well as humans, it is generally agreed that such states are basically aversive.

These descriptions of affective defense behavior in humans bear a striking resemblance to the term "episodic dyscontrol" described in detail by Monroe, (1978). According to Monroe, episodic dyscontrol (or intermittent explosive disorder) is a term that characterizes an "explosive personality", reflecting the absence of impulse control. This behavioral condition has often been associated with paranoia, and altered perceptual states and presumably occurs in response to stimuli that evoke fear, anger, or rage. Episodic dyscontrol assumes the presence of excessive neuronal discharges from limbic structures to subcortical regions such as the hypothalamus and brainstem. In fact, this represents a central thesis of Monroe who provided extensive evidence in support of this view.

Predatory attack

Predatory attack, which is common among a wide variety of species, has as its major objective the procurement of food for the aggressor, and for this reason, occurs across species. I also pointed out that few autonomic signs are present in this form of aggression aside from some mild pupillary dilatation. The response pattern is not associated with aversive properties, but is, instead, positively reinforcing. The question of importance here is whether or not a comparable form of behavior exists in humans. This question is addressed directly in the following discussion.

Perhaps, the most extensive description of human predatory aggression was provided by Meloy [1988, 1997]. Of central importance is of this description that the characteristics of human predatory aggression contrast with those associated with affective defense. In particular, Meloy points out the virtual absence of sympathetic signs that are characteristic of affective defense. Because of the absence of these signs, it is often difficult to detect any response patterns that could be used to predict the onset of predatory aggression. It should also be noted that individuals displaying predatory violence might shift to affective aggression when the victim is in physical contact with the aggressor. The trigger for the shift in response patterns is the physical presence of the victim, which likely activates acute anxiety, fear or anger reactions. It is further possible that the reverse sequence may take place, namely, that predatory aggression may follow affective aggression as a means of causing more punishment to the victim. This behavioral pattern may be particularly true with psychopaths who express sadistic impulses [Meloy, 1988, 1997].

A second characteristic of predatory aggression is that there appears to be little conscious awareness of emotion. If there is any emotion at all, it is one associated with positive reinforcement, in which case, the individual may have possessed feelings such as exhilaration. Accordingly, this form of behavior may be viewed as sociopathic in nature. The aggressive act

will also heighten self-esteem, resulting in greater sense of self-confidence and sadistic pleasure. Such feelings contrast dramatically with affective defense, which is associated with aversive feelings.

A third property is that, similar to the cat, the behavior is purposeful and planned. The attack is purposeful in that the aggressor chooses the target, the manner of attack, and the magnitude of the response. In this way, it parallels predatory attack in subhuman species. One major difference, however, is that in subhuman species such as the cat, predatory attack is typically directed against an animal of another species. In humans, the attack is directed against other humans. The exception here would be the human "sport" of hunting, which of course is directed against lower species. Concerning the motivation underlying such behaviors, in animals such as felines, the purpose is one of food-seeking behavior. But, what is the motivation in humans? Certainly, it would seem that food-seeking behavior should play little or no role since food is readily available in supermarkets or local grocery stores in most civilized societies. Meloy suggests, instead, that predatory behavior "may be used to gratify certain vengeful or retributive fantasies. Or, it may reflect the behavior of a "hit man" who has no knowledge or feelings to the target individual but whose reward is purely financial.

A fourth property is that there is no perceived threat. Instead, the aggressor rather than his responding to a threat by an opponent, which occurs during affective defense, actively seeks the target. Of interest here is that the aggressor's active approach to the target can be considered a form of "stalking" which may represent an homologous form of behavior to that elicited by the cat in its "stalking" of a prey object. A fifth property is that predatory aggression may be triggered by a variety of objectives such as gratification of sadistic desires and fantasies, and relief from compulsive drives. This contrasts with affective defense where there is a single objective of reducing the perceived threat.

Regions and pathways mediating defensive rage and predatory attack

A principal goal in the study of the neurobiology of aggression and rage is to identify the underlying mechanisms of the respective behaviors. This objective requires knowledge of the anatomical substrates of these behaviors. The following discussion identifies the sites within the hypothalamus and midbrain from which each of these forms of aggression are elicited and the pathways mediating these behaviors to other regions of the brainstem and related areas of the central nervous system.

Affective defense behavior

Defensive rage behavior can be elicited by electrical stimulation of wide regions along the rostro-caudal axis of the medial hypothalamus. Defensive rage is also elicited by electrical or chemical (i.e., glutamate analog) stimulation of the dorsolateral quadrant of mainly the rostral half of the PAG. Components or fragments of defensive rage can also be elicited from

lower regions of the brainstem. These regions include the caudal PAG and pontine tegmentum and presumably lie along the descending pathways mediating this form of aggression.

The principal descending pathway from the medial hypothalamus subserving defensive rage behavior arises from the anterior medial hypothalamus and its primary target is the dorsolateral aspect of the rostral half of the PAG. The functions of this pathway are mediated by glutamate acting upon NMDA receptors in the PAG. Of particular interest is that other regions of the medial hypothalamus, such as the ventromedial nucleus from which defensive rage can also be elicited, project rostrally to the region of the anterior medial hypothalamus from which the descending pathway to the PAG arises. Moreover, the anterior medial hypothalamus also receives significant inputs from components of the limbic system (described below) which modulate aggression and rage behavior. The converging inputs into the anterior medial hypothalamus thus enables this region to serve as a major site of integration for the expression of defensive rage behavior.

The second limb of the descending pathway for the expression of defensive rage behavior arises from the region of the dorsolateral PAG, which receives direct inputs from the anterior medial hypothalamus. The efferent projections of this region of the PAG are directed to structures that mediate autonomic and somatomotor components of defensive rage behavior. There are several routes by which autonomic functions are activated from the PAG. One pathway includes a projection to the locus ceruleus, which in turn projects to the intermediolateral cell column of the thoracic and lumbar spinal cord. Converging inputs to these sympathetic regions of spinal cord are also mediated through projections to the solitary nucleus, whose axons then project to the ventrolateral medulla and from there to the intermediolateral cell column of the thoracic and lumbar cord. There are several regions that mediate the somatomotor components of defensive rage behavior. One set of targets includes the motor nuclei of the trigeminal and facial cranial nerves, which are associated with jaw opening essential for the vocalization aspect of the defensive rage response. A second target includes the nuclei of the reticular formation which comprise, in part, reticulospinal fibers directed towards alpha and gamma motor neurons. Those neurons directed to the cervical cord presumably affect movements of the upper limbs that comprise the striking component of the rage response. It is the collective integration of these two components that are integrated at the levels of the medial hypothalamus and PAG, which comprise the defensive rage response. A separate, ascending projection of the dorsolateral PAG supplies the rostro-caudal extent of the medial hypothalamus, much of which relate to the expression of defensive rage. This projection likely serves as a substrate for a positive feedback mechanism, thus increasing the likelihood that this response can be prolonged under dangerous conditions, which is of survival value to the animal.

Predatory attack behavior

Predatory attack behavior can be elicited by electrical stimulation most typically of the perifornical lateral hypothalamus, ventrolateral aspect of the PAG, and ventral tegmental area.

The principal origin of the descending projections of the hypothalamus is the region of the perifornical lateral hypothalamus from which predatory attack is elicited. This region supplies the ventrolateral aspect of the PAG, ventral tegmental area, central tegmental fields of the midbrain and pons, locus ceruleus, and motor and main sensory nuclei of the trigeminal complex. The projections to the trigeminal complex are significant in that they provide the anatomical substrate for the jaw closing reflex critical for the culmination of biting attack. The projections to the brainstem tegmentum presumably provide the initial neuron in a series of descending projections to the lower brainstem and spinal cord essential for other motor aspects of the attack response such as stalking and striking at the prey object.

Anatomical and functional relationship between the medial and lateral hypothalamus

While defensive rage behavior and predatory attack clearly reflect distinctly different forms of aggression that utilize separate and non-overlapping pathways, they also relate to each other in a unique manner. Within the medial hypothalamus and with respect to defensive rage behavior, there are two classes of neurons. One is a projection neuron, which was described above, whose target is the dorsolateral aspect of the PAG and constitutes the descending pathway for this form of aggression. The second is a neuron with a short axon that supplies the lateral hypothalamus. It is GABAergic and inhibits neurons in the lateral hypothalamus associated with predatory attack. Likewise, there are at least two classes of neurons in the lateral hypothalamus with respect to predatory attack. The first is a neuron with a long axon that constitutes the descending pathway for the expression of predatory attack and the second is a GABAergic neuron, with a short axon which supplies the medial hypothalamus and inhibits neurons in the medial hypothalamus associated with defensive rage. The likely functional significance of the reciprocal inhibitory pathways linking the medial and lateral hypothalamus is as follows: since these two responses are mutually exclusive, the effective expression of one requires the suppression of the other. It is intuitive that a successful act of predation can only be accomplished when a predator quietly approaches the prey object, which requires suppression of hissing and related component responses. Similarly, when defensive rage is required following the presence of a threatening stimulus, elements of predation serve no function and therefore are suppressed in order for the affective components of the response to become manifest. Collectively, the neuroanatomical relationships between the medial and lateral hypothalamus thus provide the essential substrates which are of survival value to the animal.

Limbic structures associated pathways modulating aggression and rage

The limbic system consists of the following structures: amygdala, hippocampal formation, septal area, prefrontal cortex and cingulate gyrus. They possess several common anatomical and functional features which distinguishes them from other regions of the brain. These include the following: (1) they receive secondary or tertiary sensory inputs which may vary among limbic structures; (2) they receive inputs from brainstem monoaminergic neurons;

and (3) they project directly or indirectly to the hypothalamus and related structures of brainstem. These combined sensory and monoaminergic inputs serve to activate limbic structures to cause powerful modulation of aggression and rage by virtue of their efferent projections to the efferent target structures (Fig. 1). This section describes the modulating effects of limbic structures upon aggression and rage and their associated pathways over which such modulation is mediated.

Amygdala

The amygdala, which consists of a complex of nuclei located in the rostral aspect of the temporal lobe, has received more attention than any other limbic structure with respect to its relationship to emotional behavior. These studies have revealed that the amygdala is not uniform in its effects upon aggression and rage. Instead, the effects are dependent upon both the form of aggression and region of amygdala considered.

Excitation of the region of amygdala including the medial nucleus and medial aspect of the basal complex in the cat potentiates defensive rage behavior elicited from the medial hypothalamus, while excitation of the lateral and central nuclei or lateral aspect of the basal complex suppresses this response. The potentiating effects of the medial amygdala are mediated over the stria terminalis, which projects to the bed nucleus of the stria terminalis and rostral half of the medial hypothalamus, including the dorsomedial region and shell of the ventromedial nucleus. A primary neurotransmitter of this pathway has been identified and is substance P (SP), acting upon neurokinin-1 (NK_1) receptors in the medial hypothalamus. In contrast, excitation of the medial amygdala suppresses predatory attack behavior elicited from the lateral hypothalamus. Suppression is manifest via a disynaptic pathway in which the first limb includes the stria terminalis projection to the medial hypothalamus and the second a GABAergic (inhibitory) neuron projecting from the medial to lateral hypothalamus. The inhibitory effects of the amygdala upon defensive rage behavior are mediated through a descending projection to the midbrain PAG. The neurotransmitter has been shown to be enkephalin acting through μ-opioid receptors in the PAG. In a parallel manner, excitation of the lateral amygdala potentiates predatory attack. While the pathway has not been experimentally identified, it is likely to include fibers of the ventral amygdalofugal pathway projecting to the lateral hypothalamus.

Hippocampal formation and septal area

The hippocampal formation in rodents and felines is arranged in a manner that extends from its rostral tip situated in proximity to the septal area caudally, beneath the corpus callosum and parallel to the lateral ventricle, entering the temporal lobe where it passes ventrally and rostrally, ending at a position just caudal to the amygdala. The dorsal region of hippocampus (near the septal pole) suppresses predatory attack, while the ventral region (near the temporal pole) facilitates this form of aggression. As indicated below, the modulating

properties of the hippocampal formation upon aggressive behavior is mediated through the septal area, a structure which may thus be viewed as a relay nucleus of the hippocampal formation.

The pathways over which hippocampal modulation of aggressive responses are mediated likely involve the precommissural fornix—the branch of the fornix that supplies the septal area. The projection from the hippocampal formation is topographically organized in that fibers arising from the dorsal (septal) aspect project to the medial aspect of the lateral septal nucleus, while progressively more caudal regions of the hippocampal formation (toward the temporal pole) project to more progressively lateral aspects of the lateral septal nucleus. In turn, the medial aspect of the septal area, which received inputs from the dorsal hippocampal formation, projects to the medial hypothalamus. In this manner, activation of the dorsal hippocampal formation excites neurons in the medial aspect of the septal area, which in turn, excites neurons in the medial hypothalamus. Because the medial hypothalamus communicates with the lateral hypothalamus via a short GABAergic neuron, activation of the medial hypothalamus either directly or indirectly through the medial septal area or dorsal hippocampal formation causes suppression of predatory attack. Activation of the medial hypothalamus via a projection from the medial aspect of the septal area provides the anatomical basis for septal area potentiation of defensive rage behavior elicited from the medial hypothalamus.

Concerning the anatomical basis by which the ventral hippocampal formation and lateral aspect of the lateral septal nucleus potentiate predatory attack behavior, the likely pathways include a direct projection to the lateral aspect of the lateral septal nucleus, which in turn projects to and (presumably) activates neurons in the lateral hypothalamus which mediate the expression of predatory attack behavior.

Prefrontal cortex and anterior cingulate cortex

The prefrontal cortex and adjoining regions of the anterior cingulate cortex exert powerful suppression of predatory attack and rage behavior. There are a number of descending fiber systems from the prefrontal cortex that could provide the anatomical substrate for modulation of aggression. These include a monosynaptic connection consisting of small numbers of neurons that project directly to the hypothalamus from the prefrontal cortex and several multisynaptic pathways involving connections with either the amygdala or mediodorsal thalamic nucleus. Of these pathways, there is experimental evidence that the modulating effects from the prefrontal cortex and anterior cingulate gyrus upon aggression and rage are mediated primarily through the multisynaptic pathway involving the mediodorsal thalamic nucleus. With respect to this pathway, the mediodorsal thalamic nucleus projects to the hypothalamus through a series of interneurons in the midline thalamus. The neurotransmitter for this system of neurons is not known but presumably include a glutamate projection from the prefrontal cortex and anterior cingulate cortex.

References

Meloy, J.R. *The Psychopathic Mind: Origins, Dynamics, and Treatment*, Jason Aronson, Inc, Northvale, NJ, 1988, pp. 1-474.

Meloy, J.R. Predatory violence during mass murder, Journal of Forensic Science 42 (1997) 326-329.

Monroe, R.R. *Brain Dysfunction in Aggressive Criminals*, Lexington Books, Lexington, Mass., 1978, pp. 1-223.

Siegel, A. *The Neurobiology of Aggression and Rage.* CRC Press, Inc., Boca Raton, Florida, 2005, pp 1-243.

Siegel, A., Bhatt, S., Bhatt, R., and Zalcman, S. S. The neurobiological bases for development of pharmacological treatments of aggressive behavior. Current Neuropharmacology (2007) 5: 135-147.

Siegel, A., T. A.P. Roeling, T. R. Gregg and M.R. Kruk. The neuropharmacology of brain stimulation evoked aggression. Neuroscience & Biobehavioral Reviews, (1999) 23:359-389.

Siegel, A. and Victoroff, J. Understanding human aggression: new insights from neuroscience. International Journal of Law and Psychiatry (2009) 32: 209-215.

Vitiello, B., Behar, D., Hunt, J., Stoff, D., and Ricciuti, A. Subtyping aggression in children and adolescents, J.Neuropsychiatry. 2 (1990) 189-192.

Chapter 2: The Aggressive Brain

Jordan Grafman, Ph.D.
Director, Brain Injury Research
Rehabilitation Institute of Chicago
Chicago, Illinois
&
Pamela Blake, M.D.
Memorial Hermann Northwest Hospital
Houston, Texas

Introduction

Humans, like other animals, start out aggressive. Toddlers and small children often use aggressive actions in response to environmental stimuli. During maturation, however, the developing brain, in concert with societal influences, begins to exert an inhibitory effect on our innate aggressive tendencies, so that by the time we are adults, aggressive actions are frowned upon in most modern societies. Even in generally non-violent societies, however, under certain circumstances, such as sport (e.g., boxing or football) or warfare or self-defense, aggression is permitted or even sanctioned. As we enter late adulthood, degenerative brain conditions such as Alzheimer's Disease, with their attendant loss of the brain areas that exert an inhibitory influence on aggression, can create a tendency for aggressive behaviors to reappear.

Some people appear predisposed to being more aggressive, and, in extreme cases, criminal violence may run in families. Certain cultures and environments also permit aggressive and violent behavior and thus both genetic predisposition and environmental exposure can contribute to a tendency to being more aggressive. Even in violent societies that permit frequent aggressive behavior, "accepted" aggression falls within the boundaries of certain agreed upon rules of behavior and individuals who break those rules are often singled out as being excessively violent or aberrant. Inevitably, the neurosciences are being recruited to help explain the brain basis of aggression and violence and to perhaps offer some clues to the modulation of such behavior (Davidson, Putnam, & Larson, 2000).

Different patterns of aggression have been identified, and can be divided into impulsive, or reactive, aggression, and premeditated, or planned, aggression. Impulsive aggression is more often encountered in society. It involves an unplanned, inappropriate act of aggression in response to an environmental trigger and can be seen in behaviors such as road rage, spontaneous bullying, or attacking another person in response to a verbal provocation. Impulsive aggression is the type most often encountered with children and adolescents when the brain is still maturing and tends to become less frequent when a person reaches adulthood. Premeditated aggression, on the other hand, is less commonly encountered but potentially more deadly and more dangerous. It is typically committed by the mature individual and can involve elaborate degrees of planning. Examples of premeditated aggression include serial killers, assassins, and terrorists. As we will see below, since impulsive aggression is a 'normal' neurological function in the developing brain that later becomes latent, the conditions that trigger spontaneous aggressive tendencies can be created in a laboratory setting to allow the study of brain function related to impulsive aggression. Premeditated aggression does not lend itself as easily to formal study and thus much less is known about the neurobiology of this condition.

Clinical neuroscientists have taken a number of different approaches to understand the contribution of various brain regions to the exhibition of aggression and its management. In this brief review, we will focus on our laboratory's work examining the key brain regions involved in impulsive aggressive behavior and control, the types of environmental exposure that could influence aggressive behavior, and the interaction of genetic predisposition, brain damage, and aggressive behavior.

Key Brain Regions Involved in Aggressive Behavior

When we began the Vietnam Head Injury Study (VHIS) in the early 1980's (Raymont, Salazar, Krueger, & Grafman, 2011) the neurobehavioral evaluation of patients was focused on

the cognitive and personality consequences of their penetrating brain injuries. Within the first year of the study, however, we began to receive letters from caregivers, mostly spouses, indicating that we weren't assessing some of the key problems they were experiencing at home including an increase in abnormal social behaviors often characterized by impulsive aggression. Enlisting the help of family members of the veterans participating in the VHIS (almost always a spouse), we asked them to complete several scales and forms measuring aggressive behavior and the functioning of our participants. We then constructed a violence and aggression scale and examined the association between the degree of aggression reported by the spouse, the daily functional performance of participants who had high or low aggression scale scores, and the location of their brain damage (Grafman et al., 1996).The location of their brain damage that was most closely associated with increased aggression was the ventromedial prefrontal cortex (vmPFC), which is located in the bottom part of the frontal lobes of the brain. The frontal lobes are located behind the forehead, and the ventral (bottom part) is located just above the eyes. The medial part is in the middle. The frontal lobe of the brain is more developed in more advanced species of animals, and is known to be important for higher cognitive functions such as organization, planning, control of impulses, and working memory. In the Vietnam Head Injury Study, those patients who had vmPFC lesions were significantly more aggressive than patients with lesions in other brain areas or matched non-injured Vietnam combat veterans. Our interpretation was that the vmPFC was critical for storing the social regulatory rules that we learn during development, and that these regulatory rules inhibit more primitive aggressive behaviors. When that area is damaged, the ability to access those rules is degraded, loosening the inhibitory reins on aggression, and leading to more aggression. Although the aggression reported by the patient's significant other was primarily anger, yelling or throwing things rather than physical violence and criminal activity (although these latter more violent behaviors did occur in a small proportion of the subjects we studied), this abnormal degree of aggression was extraordinarily disruptive to the family, leading to a disintegration of family relations and responsibilities. This relatively large-scale lesion study confirmed the relevance of the vmPFC region of the frontal lobes for the self-control of aggressive behavior. Consistent with this theory, studies of individuals who have been aggressive, usually incarcerated for violent acts such as assault or homicide, have shown abnormalities present on the neurological examination that indicate impairment of frontal lobe function (Blake & Grafman, 2004). We note, however, that not all people with vmPFC lesions demonstrate increased aggression, and that people can be aggressive without having vmPFC damage, so other factors clearly play an important role in the expression of aggression.

Genetics and Aggression

One long-standing area of research is the relationship of aggressive behavior to genetic predisposition. Although a number of genes have been implicated in aggressive and violent behavior, perhaps the most studied gene is the MAO (monoamine oxidase) gene and its polymorphisms. A genetic polymorphism indicates that a single gene can have normal variations in genetic structure that can sometimes lead to differences in behavior or function. We examined the effects of having one of two variants of the MAO gene in Vietnam Veterans with penetrating brain injury on aggressive behavior using the relative's version of the Neuropsychiatric Inventory (NPI) to record aggressive behavior (Pardini et al., 2011). Higher NPI scores were linked to more self-reported severe childhood psychological traumatic experiences and post-traumatic stress in both a control group of Vietnam Veterans without brain damage and in the group with non-frontal injury. Ventral prefrontal cortex lesions resulted in more aggression than lesions elsewhere in the brain. The combat veteran controls without brain damage that had a particular version of the MAO gene tended to be more aggressive, replicating the literature. Patients with damage to posterior brain regions also showed the same effect. Patients with ventral frontal lobe lesions did not show a polymorphism effect. This indicates that polymorphism effects upon behavior are dependent upon the intactness of brain regions mediating the behavior of interest. When such a region is damaged (as in the case of the ventromedial prefrontal cortex), then the polymorphism should have a minimal effect upon the targeted behavior; in this case, aggression. This finding reaffirms the importance of the vmPFC for mediation of aggression-controlling behavior. In a follow up study, Pardini et al (submitted) examined the effects of the DRD1 dopamine receptor gene and the COMT gene on rated aggression using the NPI in the same sample of Vietnam Veterans with brain damage. The results indicated that patients who had a particular variant of the DRD1 gene compared to non-carriers of that version demonstrated greater aggression after medial lesions in the frontal lobes but reduced aggression after lateral frontal lobe lesions. A COMT gene variant had no effect on reported aggressive behavior indicating that only selected genes (even among those that get expressed in the frontal lobes) can influence aggressive behavior.

Influence of Environmental Exposure to Aggression

So far we have reported that specific brain lesions can affect impulsive aggressive behavior, and that the ventral prefrontal cortex is a key brain site modulating impulsive aggressive behavior (Blake & Grafman, 2006). Next we found that genetic predisposition can have an effect upon the expression of aggression as long as the key brain area mediating that effect is not damaged.

We also found that childhood trauma and stress (as self-reported in a questionnaire) predicted aggressive behavior in adults with an intact frontal lobe. This finding highlights how important environmental influences are upon the expression of aggression (Tuvblad & Baker, 2011). To examine this relationship in more detail, we decided to study adults and adolescents who were healthy and exposed to aggression and violence through video games, other media, imagination, or in real life. None of the subjects volunteering for these studies had any reported history of inappropriate aggression. In the studies with adolescents, we obtained psychophysiological measures and functional magnetic resonance imaging while they imagined being aggressive or a victim of aggression or were simply exposed to aggressive media. Adults were studied only during a simulated aggressive act. We also evaluated the participants' real-life exposure to aggressive media as well as their cognitive and personality characteristics. While it is hard to study true impulsive or planned violent behavior in the laboratory, we can try to observe the brain in action when someone is imagining an unplanned aggressive act in response to a request by the experimenter.

In the first study, we asked adults to imagine being in an elevator with their mother who was subsequently, in the experimental conditions, assaulted by someone entering the elevator (Pietrini, Guazzelli, Basso, Jaffe, & Grafman, 2000). The subject was asked to respond in a number of ways: in one condition, he imagined himself being restrained and unable to respond, in another condition, he simply didn't respond, or in the key condition, he became aggressive to defend his mother. As the subject imagined becoming more aggressive across conditions, increasingly diminished activity was observed in ventral prefrontal cortex. We interpreted this finding to indicate that as a person becomes more aggressive, it is necessary to suppress activity in brain areas concerned with instantiating social rules that inhibited aggression. If one wants to fight effectively, social rules against fighting must not be prominently activated in the brain so that the brain regions that mediate aggression can operate without interference.

We wondered if we would see the same effect in adolescents in a similar task involving the theft of their jacket from an underground parking lot (Strenziok, Krueger, Heinecke, et al., 2011). Once again we found reduced ventromedial prefrontal cortex (vmPFC) activation associated with imagined aggressive behavior as well as *enhanced* aggression-related activation in the frontopolar cortex (FPC) but only with increasing age. The enhanced activation in the FPC was also associated with a normal developmental structural change in brain tissue density – a thinning of that cortical area of the PFC. This increase in FPC activation was associated with judgments of the severity of aggressive acts. Reduced vmPFC activation was associated with greater aggression reinforcing that its normal function is to exert inhibitory control over aggressive impulses. Concurrent FPC activation likely reflects foresight of harmful consequences that result from aggressive acts. The correlation of age-dependent functional activation

changes and structural cortical thinning demonstrates ongoing maturation of the FPC during adolescence towards a refinement of social and cognitive information processing that can potentially lead to non-aggressive responses when provoked.

Next, we examined whether there is a relationship between exposure to violence and brain structural development and functional activity. First we examined the relationship between cortical grey matter density and media violence exposure in healthy male adolescents using a technique that measures the density of cerebral cortex along with a questionnaire measuring self-reported exposure to aggression and violence (Strenziok et al., 2010). Adolescents with more frequent exposure to aggression and violence, (usually due entirely to mass media or gaming exposure), had lower left lateral orbitofrontal cortex density--a possible risk factor for altered socioemotional functioning. Many surveys have indicated that adolescents spend a significant part of their leisure time playing with media that includes aggressive activities or watching TV programs and movies that portray violence. Although we found a relationship between brain structure and exposure to violent media, it is unknown how the extent of violent media use and the severity of aggression displayed affects adolescents' brain function. Thus, we decided to investigate the relationship of skin conductance responses (a psychophysiological measure of stress and emotion), brain activation and functional brain connectivity to concurrent media violence exposure in healthy adolescents (Strenziok, Krueger, Deshpande, et al., 2011). In an event-related functional magnetic resonance imaging experiment, adolescents repeatedly viewed normed videos that displayed different degrees of aggressive behavior. We found that skin conductance responses decreased dramatically with increasing aggression suggesting that our adolescent subjects were becoming desensitized emotionally after they watched these aggressive scenes. Our results further revealed similar brain desensitization that resulted in lower brain activity in social-emotional brain network including the frontal lobes. Further analyses revealed the particular importance of the left ventral frontal lobes, and also showed that brain activation during viewing aggressive media decreased *over time*, for more aggressive videos. We concluded that aggressive media inhibits an emotion-attention brain network that has the capability to blunt emotional responses through reduced attention with repeated viewing of aggressive media contents. We believe that the brain changes just described restrict the linkage of the consequences of aggression with an emotional response, thereby inducing the tolerance of aggressive attitudes and behavior.

Although in our studies of people with brain damage or normal brain functioning have emphasized the importance of the frontal lobes for the routine inhibition of inappropriate aggressive behavior, areas in the anterior temporal lobes such as the amygdala (concerned with the emotional labeling of behavioral acts or postures) and superior anterior temporal cortex

(concerned with the semantic storage of social behaviors independent from the context in which those behaviors are observed or emitted) are also involved (Matthies et al., 2012; Zahn et al., 2007), particularly with pathologic violence such as that committed by psychopaths in whom fear conditioning is pathologically and chronically decreased (Coccaro, McCloskey, Fitzgerald, & Phan, 2007). Such pathological changes in fear conditioning have been observed in children who later met diagnostic criteria for psychopathy.

Can we use this kind of data to predict who might be aggressive? We imagine that clinical researchers could develop a reasonably accurate prediction in individuals willing to undergo a battery of tasks that provoked impulsive or planned aggression, completed forms indicating past aggression exposure or activity (with verification by significant others), reported the current environmental context that they find themselves in, allowed genetic information to be gathered, and agreed to be tested on measures of personality, inhibition, executive function and social cognition. Acquiring all this information with permission from any individual would be challenging enough. But even with the accumulation of all of this information, there is no guarantee of identifying with anywhere near 100% accuracy people likely to commit an impulsive or planned violent act. One could, perhaps, construct a sensitive test battery with little specificity since it would be likely that most of the people identified through this provocative and predictive evaluation tool would not be killers or excessively violent.

Although few people turn out to be capable of committing violent crime or dangerously aggressive acts, everyone has an opinion about that kind of aggression and violence and what the consequences of such behavior should be. In that latter case, we have determined that repeated exposure to violence or aggression desensitizes people which should make them become more tolerant of aggression and violence by diminishing the emotional and even social response to aggression in others (or potentially themselves). This would have the consequence of creating a more permissive society regarding aggression. If society is more accepting of aggression, then that would change the norms of the culture and potentially promote aggressive behavior in at-risk individuals who under different social rules might be more inclined to inhibit aggressive tendencies. Thus, by modifying the environment, and therefore the context, in which aggression occurs, it may be possible to reduce or increase, overall, the likelihood of aggressive acts. To enact aggression reduction measures would mean controlling exposure to provocative and aggressive media, punishing aggressive acts forcefully with due process, educating the public about the negative consequences of aggression, bullying, and violence outside of sanctioned activities (and promoting economic policies like increasing gainful employment that tend to reduce crime). In our view, the context in which aggression and violence occur can be modified much more easily than identifying individuals likely to

commit aggressive acts. By manipulating context, society can reduce aggression by individuals indirectly.

Our studies imply that the earlier in life that exposure to inappropriate violence is limited, the more likely that this modification will reduce the frequency of inappropriate aggression and its acceptance in adulthood (Down, Willner, Watts, & Griffiths, 2011). This early intervention would also presumably promote the enhancement of brain regions designed to inhibit inappropriate behavior in individuals. But the focus here is on modifying societies to prevent future violence, rather than identifying individuals at risk for committing that violent behavior. Although we can determine with fair accuracy individuals who have psychopathic tendencies, often this determination of psychopathy in violent individuals is after the fact. There may still be ways (see above for the kind of evaluation that would be needed) to identify some individuals most at risk for aggressive and violent behavior and to intervene in those cases. There is evidence that cognitive-behavioral therapy is modestly effective in reducing aggressive behavior in individuals (Down, et al., 2011). Finally, we have to determine what is the best solution to restricting violent media exposure in children. As the epicenter of violent behavior media has shifted to the web and hand-held portable devices from television and movies and even game player stations, it will be a challenge to filter such aggressive content in a developmentally appropriate way without portraying the world in a manner inconsistent with a child's experience or knowledge or appearing as if it is unduly censoring content that should be permitted in a free society-even for children.

In summary, even imagining aggression diminishes activity in the human frontal lobes, a region concerned with inhibiting inappropriate behavior and supporting appropriate social conduct. We would predict that a similar if not more severe reduction in frontal lobe activity occurs during actual aggression. Certain individuals have genetic polymorphisms that predispose them to being more aggressive. Repeated exposure to aggressive acts also diminishes activity in crucial brain areas concerned with social reasoning and emotion thereby gradually disconnecting the aggressive activity from its consequences. None of these findings can aid in the identification of a specific person who is guaranteed to commit an aggressive act sometime in the future without evidence of that person having committed such an act previously. So then how can the studies we conducted be used to understand aggressive behavior and violence? One obvious finding is that selected brain areas and genes are important for the inhibition of inappropriate aggressive activity and this knowledge could indicate brain areas to target in various kinds of interventions from behavioral to pharmacologic. Our findings also suggest that cultural exposure to violence would have a role in modulating the emotional and reasoned response to violence and point to the importance of modifying the culture the targeted population is in to reduce acceptance of inappropriate

aggressive behavior. Finally, it is necessary to continue to study new tools or sets of measures that will likely improve the accuracy of identifying individuals likely to commit an inappropriate or criminal aggressive act. The implications of all of this research for society are obvious.

References

Blake, P., & Grafman, J. (2004). The neurobiology of aggression. *Lancet, 364 Suppl 1*, s12-13.

Blake, P., & Grafman, J. (2006). Effect of orbitofrontal lesions on mood and aggression. In D. H. Zald & S. L. Rauch (Eds.), *The Orbitofrontal Cortex* (pp. 579-595). Oxford: Oxford University Press.

Coccaro, E. F., McCloskey, M. S., Fitzgerald, D. A., & Phan, K. L. (2007). Amygdala and orbitofrontal reactivity to social threat in individuals with impulsive aggression. *Biological psychiatry, 62*(2), 168-178.

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*(5479), 591-594.

Down, R., Willner, P., Watts, L., & Griffiths, J. (2011). Anger Management groups for adolescents: a mixed-methods study of efficacy and treatment preferences.*Clinical child psychology and psychiatry, 16*(1), 33-52.

Grafman, J., Schwab, K., Warden, D., Pridgen, A., Brown, H. R., & Salazar, A. M. (1996). Frontal lobe injuries, violence, and aggression: a report of the Vietnam Head Injury Study. *Neurology, 46*(5), 1231-1238.

Matthies, S., Rusch, N., Weber, M., Lieb, K., Philipsen, A., Tuescher, O., et al. (2012). Small amygdala-high aggression? The role of the amygdala in modulating aggression in healthy subjects. *The world journal of biological psychiatry : the official journal of the World Federation of Societies of Biological Psychiatry, 13*(1), 75-81.

Pardini, M., Krueger, F., Hodgkinson, C., Raymont, V., Ferrier, C., Goldman, D., et al. (2011). Prefrontal cortex lesions and MAO-A modulate aggression in penetrating traumatic brain injury. *Neurology, 76*(12), 1038-1045.

Pietrini, P., Guazzelli, M., Basso, G., Jaffe, K., & Grafman, J. (2000). Neural correlates of imaginal aggressive behavior assessed by positron emission tomography in healthy subjects. *The American journal of psychiatry, 157*(11), 1772-1781.

Raymont, V., Salazar, A. M., Krueger, F., & Grafman, J. (2011). "Studying injured minds" - the Vietnam head injury study and 40 years of brain injury research. *Frontiers in neurology, 2*, 15.

Strenziok, M., Krueger, F., Deshpande, G., Lenroot, R. K., van der Meer, E., & Grafman, J. (2011). Fronto-parietal regulation of media violence exposure in adolescents: a multi-method study. *Social cognitive and affective neuroscience, 6*(5), 537-547.

Strenziok, M., Krueger, F., Heinecke, A., Lenroot, R. K., Knutson, K. M., van der Meer, E., et al. (2011). Developmental effects of aggressive behavior in male adolescents assessed with structural and functional brain imaging. *Social cognitive and affective neuroscience, 6*(1), 2-11.

Strenziok, M., Krueger, F., Pulaski, S. J., Openshaw, A. E., Zamboni, G., van der Meer, E., et al. (2010). Lower lateral orbitofrontal cortex density associated with more frequent exposure to television

and movie violence in male adolescents. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine, 46*(6), 607-609.

Tuvblad, C., & Baker, L. A. (2011). Human aggression across the lifespan: genetic propensities and environmental moderators. *Advances in genetics, 75*, 171-214.

Zahn, R., Moll, J., Krueger, F., Huey, E. D., Garrido, G., & Grafman, J. (2007). Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 104*(15), 6430-6435.

Chapter 3: Role of Genes/environment

Peter K. Hatemi, Pennsylvania State University, phatemi@gmail.com
Rose McDermott, Brown University, Rose_McDermott@Brown.edu

Much aggression is motivated by conflict between in-groups and out groups. What incites the desire to aggress, as opposed to seeking more conciliatory mechanisms of negotiation? How do we gauge what type of stimuli will either build up or dissuade the tendency to respond to threats aggressively among certain types of people? Different media strategies, from the use of videos, to particular framing of certain events, may either heighten or reduce prospects for people under threat or stress to respond aggressively to provocation. Learning more about that process, and the triggers that enhance or diminish prospects for such reaction, may prove particularly beneficial to reduce the number of recruits into violent extremist organizations, both at home and abroad. Such knowledge may help us build better tools to reduce the number of susceptible individuals, or to provide truly effective counter-messaging strategies that can provide alternative stimuli that could intervene in the pathway from provocation to violence.

So far, most research in this area has focused on the environment and specific stimuli. However, understanding how biological and genetic mechanisms contribute to aggression is critically important and holds myriad real world implications. Indeed, conditions, messages and other stimuli must be processed in order for them to have an effect on behavior. Though the greater pathways of cognition and emotion are universal in humans, there is wide variance on how individuals process their conditions, and become motivated to act upon them. These differences are both social and biological. Here we review some of the potential mechanisms by which genetics and physiology inform one's psychological architecture and contribute to decision making, particularly around aggression. The role of neurobiology in aggression might appear obvious to clinicians or those who research the topic, but it may appear, at first, more remote to political decision makers and field commanders. Therefore, we discuss the import of this approach in a non-technical manner in the hopes the information provided might begin a larger conversation with those who may most benefit from inclusion of neurobiological factors in their decision calculus.

Genetic influences contribute indirectly to all human experience and behavior. Yet it is not the case that a specific genotype, or hormone, will produce a particular outcome, or create a predictable reaction. Rather, thousands of genes operate in constant reciprocal interaction with myriad environmental cues, triggers and forces, regulating hormonal states that produce the diverse psychological, emotive and cognitive states that result in numerous social behaviors, including violence. The general processes appear universal in humans, but genetic dispositions that account for sensitivity to the environment vary across individuals. By taking into account both genetic and environmental characteristics, we can better elucidate the pathway toward aggression, and begin to explain how various environmental factors, such as the media and propaganda, as well as more informal mechanisms of narrative messaging, can be used by ourselves, allies and adversaries alike to manipulate the neurobiological mechanisms that inform the psychological architecture of susceptible individuals. This includes ascertaining the probabilities that one might select into groups and peers that lead to violence.

Several critical features help define how *neurobiological* models may be specifically applied to help explain aggression: 1) certain stimuli instigate general processes of emotion universal to the population, leading to aggressive behaviors. In order to fully understand how to manipulate these processes or diffuse their manipulation by others, a neurobiologically informed strategy is needed; 2) specific high-risk environments/stimuli have a stronger effect on dispositionally (e.g., genetically) sensitive individuals; and 3) individuals tend to self-select into environments that reinforce their specific vulnerability (e.g., "genotype-environment correlation"). A neurobiological approach locates causality at the intersection of general processes of cognition and emotion, and an individuals' unique disposition and their specific social contexts (e.g., "genotype x environment interaction"). That is, aggression and violence cannot be understood if we ignore basic neurobiological processes universal to humans or individual differences between people embedded within specific cultures; the roots of political violence are multifactorial, resulting from interactions between large numbers of biological (genetic) and social (environmental) factors, and these interactive effects may differ profoundly both within and across populations.

Behavioral genetic analyses suggest that somewhere between 44%–72% of the variance in aggression is due to genetic influence (Miles and Carey 1997). However, while such models display these influences as static, in reality they only hint at the role of gene-environment interplay in observing, experiencing and taking part in aggression. We highlight several of the best-explored biological mechanisms (associated genotypes and hormones) that have a substantial role in aggression and vary across individuals. First, serotonin has a major role in prefrontal cortical activity (brain), including the Orbital Frontal Cortex and Anterior Cingulate Cortex, both of which are involved in regulating aggression. Deficiencies in serotonergic innervation remove the inhibitory processes on aggression (Wood et al. 2006). De Boer et al. (2009) provide one of the most compelling illustrations of the importance of serotonin on the emergence of violence when provoked. They demonstrate how violence "trains" genetic expression, which then supports more violence. De Boer engineered mice with specific kinds of serotonin receptors and found that constitutionally aggressive individuals develop gradually over the course of repetitive exposures to victorious social conflicts involving offensive

aggression. However, this effect was prominent only through multiple victorious encounters. That is, brain serotonin activity decreases as a *consequence* of acquiring repeated victorious experiences and adopting unchecked forms of aggression. Caramaschi et al., (2008) further documented that only *after* repeated resident-intruder fighting experiences were serotonin levels in the prefrontal cortex found to be significantly lower among highly aggressive mice. These neurobiological pathways are rewarded by victory in combat and influence certain individuals to behave even more aggressively once provoked. Such a model could be applied to human aggression and war, suggesting that those who engage in more successful aggressive fighting, as might happen over multiple tours of deployment for example, or as the result of accomplishing successful terrorist attacks, will prove more likely to further escalate their acts of aggression. It also provides a model of reward that develops as a result of aggression. That is, the reason for first engaging in violence may differ remarkably from continued acts of violence. It also proffers some reason as to why it becomes difficult to diffuse violent actors even when the initial reason for violence has passed.

The low activity forms of the Monoamine Oxidase A (MAOA) genetic polymorphism may be among the best-explored genetic mechanism believed to increase the risk for physical aggression and behavioral violence. This appears particularly true where individuals have suffered traumatic early life events (Frazzetto et al., 2007; McDermott et al., 2012) or have been provoked (McDermott et al., 2009). A four-generational study of a Dutch family, who had exhibited excessively high rates of violence, including rape, murder, assault and arson, showed that the men in this family were missing an MAO-A enzyme that breaks down norepinephrine, serotonin and dopamine. Meyer-Lindenberg et al. (2006) conducted a path breaking study that demonstrated the neurological basis for differences in violence revolved around high versus low variants on the MAOA genetic allele. They found that MAOA depletions affect processes in the brain critical to the instigation or suppression of violent action. Using brain imaging, they showed that individuals with the low activity form of the allele previously associated with instances of impulsive violence had less volume in the part of their brain typically associated with emotion, the amygdala, and yet this area became hyper-responsive during emotional arousal. Concomitantly, these subjects displayed simultaneous reductions in the part of the brain associated with logical and rational decision-making in the pre-frontal cortex. Together, these findings suggest that men who have reduced emotional regulation and less cognitive control over their responses have the low activity form of MAOA.

More recently, the role of vasopressin and oxytocin have been implicated in aggression. Vasopressin and oxytocin are part of the hypothalamo-neurohypophysial system and are related to parental behavior, territorial aggression, mating and affiliation (Gordon et al. 2008). Vasopressin concentrations are significantly related to an increase in a life history of aggression and higher densities of vasopressin were associated with greater selective aggression toward unfamiliar others (Gobrogge et al. 2007). In animal knock-out studies, mice engineered without the Vasopressin 1b receptor (genetic variant) exhibited decreased levels of this type of aggression (Wersinger et al 2007). The second hormone in this group, oxytocin is best known for its importance in trust and affiliative behavior (Kosfeld et al., 2005). Lower amounts of oxytocin are believed to contribute to increased levels of fear and mistrust, which leads to

aggression. This is exactly what has been found with oxytocin knock-out mice, who also display intensified aggression. These hormones do not act in isolation. Typically when territory or kinships are threatened, aggressive behaviors become manifest; differences in these genotypes and hormones then become critically important in differential levels of aggression.

Finally, the dopamine system, in part, regulates and produces certain neurochemicals in the brain that modulates risk/reward, impulsivity and novelty seeking systems. If the dopaminergic system underlies the human experience of reward, which in its purest form encourages people to survive by seeking food, water, temperature regulation, sex, sleep, protection, or the destruction of one's "enemies", then understanding how various messages tap into this system, once other "defensive" systems are triggered, might provide one avenue by which to understand the neurobiological import of message content. This, in turn, can help identify how the specific framing of messages might either trigger aggression or help develop messages that provide people with a positive sense of reward, purpose or meaning in their lives.

The above only represent a handful of possibilities, yet the potential points of interaction and application to political violence appear evident. The probability of an individual perceiving acts of occupation or military intervention as threat to their way of life might be universal, but how individuals respond to such threats are in part regulated by dispositional differences in their genetic architecture. Understanding both the general process and individual differences may help calculate probabilities that a particular population will engage in given action. They may also be used to develop messages that may diffuse increased aggression from certain groups of individuals. That is, messages that tap into reducing anxiety from threats to one's way of life, family or territory, may have differential responses based on the complex interaction between serotonin, MAOA, dopamine, oxytocin and vasopressin systems that vary across individuals.

A great deal has been learned about impulsive, reactive and social aggression using neurobiological tools. What makes this area of great interest is that political violence in most ways represents a large scale extension of social aggression. We can model political violence as a form of social aggression instigated by actors whose neurobiological and environmental backgrounds render them differentially susceptible to respond to threat or provocation with aggression and violence. In this way, it is plausible that widespread political violence may combine elements of both reactive and instrumental aggression, or at least unify individuals who are inspired to such actions with both motivational systems. Depending on the environmental circumstances or provocations, some individuals may react to an immediate threat or oppression with violence, as when their town is being shelled, or when they have just witnessed a family member being killed. Others may engage in political violence for carefully planned strategic goals and purposes. Each of these kinds of actors may be best deterred using different incentives and restraints. Being able to distinguish between types of actors may be a key to developing effective instruments of deterrence.

Note that this perspective does not rely solely on any kind of individual genetic knowledge or manipulation, any more than it depends entirely on an environmental analysis of the messages themselves. Rather, it explores the interaction between biology in domains we know provide positive experience for people with those specific environmental cues and triggers that might either spark or diffuse aggressive responses to a variety of threats or provocations. Such information is valuable to provide probabilistic outcomes based on the populations we seek to address.

A gene by environment perspective incorporates several elements, all of which are key to understanding the susceptibility or resistance that individuals bring to bear when confronted with risk, threats, opportunities or challenges. These components begin with the genotypes that help structure individual hormonal and physiological response systems in the body. However, these systems remain permanently affected by those unique developmental and cultural features that influence people, as they grow from their time in utero through their adult experiences and relationships, until death (for full description see Hatemi and McDermott 2012). These developmental factors can include socialization experiences in the family, or in school, or personal health or social experiences that can permanently set a person on a particular path so that they are more likely to respond to threats in one way rather than other. For example, some may become fearful when confronted with threat (Hatemi et al 2012), and seek to protect themselves by placating offenders, while others instead act to defend themselves by attacking in the face of threat. These factors are critically important for calculating the probability of violence in certain populations. For example, in Romania in the 1980s, an unprecedented number of children were abandoned due to the severe economic crisis after the collapse of communism. In many orphanages, infants were left alone up to 23 hours a day, receiving almost no stimulation or human interaction. As a result, they showed severe developmental impairments in many areas of brain function, including language, learning and social attachment (Chugani et al., 2010; Eluvathingal et al. 2006). These children now experience high levels of social conflict because of their decreased capacity for empathy, among other things. Given that similar processes of social and material deprivation plague children throughout many areas of the world where famine and war are endemic, such as much of Eastern Africa, these findings help elucidate the prospects for peace or stability in the region.

As illustrated above genetic and developmental systems never exist in isolation; rather, the expression of particular genetic potential is based on critical situational contingencies, some of which occur well before the event, such as during childhood. Individuals may be capable of great violence under certain conditions but if the environment is not threatening, they may not manifest such aggressive behavior. Similarly, others seek out conditions that may serve to allow them to express their desire to commit violence. The ability of humans to be flexible and respond in a malleable manner to a variety of environmental contingencies is precisely what allows us to thrive across diverse environments over many generations.

Environmental cues are like keys of different size and shape, which can serve to unlock the specific doors they fit, each one of which releases a set of responses that calibrate to appropriate reactions, which help us distinguish food from mates, friend from foe, and threat

from opportunity. These systems may not always work perfectly, and indeed may often malfunction, but most of the time they allow most people to negotiate their way through the world successfully enough to reproduce, albeit in the midst of ubiquitous pain, loss, uncertainty and existential alienation. But, of course, certain keys may fit more than one door, and certain keys can be fabricated to elicit responses that the locksmith wants, rather than those that the homeowner may desire. Discovering the nature of the keys which unlock responses that lead to violence, and those that privilege more conciliatory responses, involves a recognition of the genotype itself (the architecture), the developmental factors which contribute to the creation of a particular phenotype (the interior decorating), but the environmental cues (keys) which entice the person to leave the house, move next door, leave town, or burn the entire place down.

Pathways for Potential Intervention

Though genetic and other neurobiological factors play a critical role in violent behavior, only environmental cues and triggers provide realistic opportunities for either turning on or turning off violent reactions. There are different points in the process that might offer opportunities for such intervention and several avenues may prove profitable to pursue. One goal may be to understand the probability of aggression based on the population characteristics, both in a neuro-developmental perspective, and from a current situational analysis. This may provide strategic information to field commanders and decision makers. Another may be more policy oriented and focused on how to diffuse violence before it begins. A third may concentrate on how to manipulate violent actors to behave more in line with U.S. interests. We have conducted research on the gene by environment interaction that shows how environmental triggers can potentiate aggressive responses (Hatemi and McDermott 2012; McDermott et al., 2009; McDermott et al., 2012). In examining a particular genotype, monoamine oxidase, we found, as have others, that the low activity form of the allele (MAOA) increases the likelihood, in an overall population, of aggressive responses. But this occurrence did not happen in isolation. Rather, this outcome only eventuated in combination with at least two other critical factors. First, the person with the particular genotype experienced some traumatic early life events. In other words, at critical developmental stages, the genetically susceptible person encountered an environment that sent signals that the world was unsafe (in our study often because the child experienced a violent early life event). The more of these problems that occurred, the more likely the MAOA susceptible child was to engage in physical aggression as an adult. Importantly, when such assaults took place mattered; puberty proved a particularly vulnerable time for individuals to encounter threats in the environment. This may be because many other internal events are also taking place simultaneously, setting down foundational patterns for life. This proves particularly important when considering it is also the recruitment age for many child-soldiers. Susceptible individuals with the low promotion genetic variant, who did not encounter traumatic early life events, were not as likely to engage in violence later in life. The second factor, which precipitated violence at higher levels, was provocation. Even those individuals who possessed the ostensible genetic susceptibility for

aggression and who had bad things happen to them as children did not wantonly engage in violence; rather, they became aggressive when they felt provoked.

Our study provides some opportunities to intervene in all three avenues described in the previous paragraph. The first is that the low promotion MAOA allele is population specific; different regions hold different allele frequencies. Understanding the basic disposition of the population might prove beneficial in calculating the potential for a violent response to US operations. A second avenue to pursue, if we seek to reduce violent action, is to target the next generation; it appears imperative to focus on addressing the children of our most egregious enemies and finding constructive alternatives for their potential for violence. Third, if we seek to manipulate current potential belligerents either towards or against aggression, we should concentrate on intervening to affect those certain environmental cues, which serve as clear triggers for aggressive behavior that might trigger the MAO, dopamine, serotonin and oxytocin and vasopressin systems we noted above.

These patterns prove particularly important for the insight they provide into avenues for potential intervention among those who possess higher genetic liability for aggression when provoked. Indeed, whether or not we know someone's genetic profile, intervening in a positive way in critical environmental arenas could benefit everyone, and reduce the prospects for violence across the board. First, environmental factors, which cause traumatic early life events, can be addressed in a more systematic way. Trivial factors are not likely to increase risk, but factors such as parental incarceration, severe parental illness or death, parental substantive abuse or depression, or child sexual or physical assault does appear to increase risk. Any one of these, as well as other likely ones such as in utero exposure to environmental toxics or maternal malnutrition, are potentially amenable to policies interventions, either abroad through USAID type of programs, or domestically at home. Second, provocation matters in potentiating violence. Foreign policies, which try to impose governments or institutions and particularly values, alien to local cultural values, are likely to be understood as constituting such provocations. Provocation lies in the eye of the beholder and what feels like help when given can be experienced as harm when received. Serious consideration should go into analyzing the perception of such policies as promotion of western values for those whose interpretation of such structures may differ from our own (see Hatemi, McDermott and Stenner 2012).

Another point in the process lies subsequent to the expression of violence itself. What can be done once a target is activated to engage in violence and aggression? Are there ways to manipulate this person back toward more constructive forms of discourse, either by breaking away from the groups that advocate violence, or by becoming an agent of positive change himself? This is where a clearer understanding of the nature of media and messaging and how such environmental factors interact with basic biology to produce the intrinsic rewards that reinforce and continue behavior in one direction or another. Our discussion of "outrage" below may prove instructive in this regard.

A Real World Problem that Necessitates Neurobiological Research: The Strategic Use of Outrage to Instigate or Motivate Violent Action

Successful leaders act as political entrepreneurs by carefully defining group members, often using arbitrary criteria to cast out-group members in a bad light, and solidifying in-group membership by appealing to the rational self-interest of those who participate and wish to receive material and status benefits from such group identification.

The process by which an individual comes to espouse a particular social or political identity occurs through a dynamic interchange between internal needs and external forces. In other words, political identity, like any other, exists, at least in part, as a function of the individual's interaction with their specific external environment, which poses its own contingencies, constraints and incentives. This allows a space for an effective leader to serve as a kind of identity entrepreneur, a person whose ability to set the agenda and define the boundaries of a particular political identity becomes a very powerful tool in mobilizing followers toward a particular partisan goal or effect. Leaders such as Bin Laden understood this concept quite well. Outrage represents a key component in consolidating support of the masses (de Toqueville 2003). Indeed, two of the most important goals in creating and maintaining political leadership reside in the ability to establish and maintain in-group solidarity while fostering out-group hostility. The use of emotional means of persuasion in service of these goals allows effective politicians or foreign agencies to essentially invoke emotional reactions on the part of citizens in order to sustain both these processes in service of a self-defined identity based cause.

An outrage occurs when a member of an opposing group, in reality or through invention or exaggeration on the part of the in-group, takes an action, or makes a statement, which members of the in-group perceive as a threat to status, by failing to take the others' values and wishes into account. By failing to show sufficient respect for the in-group, rivals in the out-group present a status challenge to in-group members, signaling that they believe they are more powerful and deserving of rights and resources than previously acknowledged or negotiated. The harm might indeed be more perceptual than real, yet such an act would signal to the opponent that the out-group not only does not offer sufficient or appropriate deference to the in-group, but also challenges the in-group to define its relative position in the status hierarchy.

The psychology that undergirds this emotional manipulation rests on biological, physiological and genetic structures. Activation of the defense mechanisms (e.g., serotonin), through threats to one's in-group and home (e.g., vasopressin and oxytocin) and the reward system to act (e.g., dopamine) become critical parts of the psychological decision process. Yet we have dedicated almost no resources to understanding these processes regarding national defense, or using these potential insights to enhance our prospects for national security. Developing strategies to more successfully manipulate environmental triggers for national security purposes and to create interventions to address the human psychological architecture, which responds to threat and outrage with aggression, can only serve to better protect US interests.

References

Caramaschi D., de Boer S. F., de Vries H., Koolhaas J. M. 2008. Development of violence in mice through repeated victory along with changes in prefrontal cortex neurochemistry. Behav. Brain Res. 3;189, 263.–272. doi: 10.1016/j.bbr.2008.01.003

Chugani, Harry, Michale Behen, Otto Muzik, Csaba Juhasz, Ferenc Nagy & Diane Chugani 2001. Local brain funcational activity following early deprivation: A stody of postinstiturionalized Romanian Orphans. *Neuroimage* 14:1290-1301.

De Boer, Sietse, Caramaschi, Doretta, Nataragan, Deepa & Koolhaas, Jaap. 2009. The Vicious Cycle Towards Violence: Focus on the Negative Feedback Mechanisms of Brain Serotonin Neurotransmission. Frontiers in Behavioral Neuroscience 3: 52.

Eluvathingal, Thomas, Harry Chugani, Michael Behen, Csaba Juhasz, Otto Muzik, Mohsin Maqbool, Diane Chugani & Malek Makki. 2006. Abnormal brain connectivity in children after early severe socioemotional deprivation: A diffusion Tensor imaging study. *Pediatrics* 117: 2093-2100.

Etheredge, Llyod. 1978. Personality Effects on American Foreign Policy, 1989-1968. *American Political Science Review* 72(2): 434-51.

Frazzetto, G. Di Lorenzo, G., Carola, V., Proietti, L., Sokolowska, E., Siracusano, A., Gross,C. Troisi, A., 2007. Early Trauma and Increased Risk for Physical Aggression during Adulthood: The Moderating Role of MAOA Genotype *PLoS ONE*. 2(5): e486. Individual Differences 51 (3): 231-236.

Gobrogge KL, Liu Y, Jia X, Wang Z. 2007. Anterior hypothalamic neural activation and neurochemical associations with aggression in pair-bonded male prairie voles. J Comp Neurol 502: 1109–1122

Hatemi, P.K., I.R. McDermott, and K. Stenner. 2012. "Reducing Recruitment into Islamic Terrorist Organizations: The Antagonistic Effect of Liberal Democracy Promotion." In *Countering Violent Extremist Organizations (Veos) Pilot Effort: Focus on Al-Qaeda in the Arabian Peninsula (Aqap)*: Strategic Command, the Chairman of the Joint Chiefs of Staff, and OSD/DDRE/RRTO, and the Office of Science and Technology Policy

Hatemi, Peter K., I.R. McDermott, M. Neale, and K. Kendler. forthcoming. "Fear Dispositions and Their Relationship to Political Preferences." *Am J Pol Sci*.

Hatemi, Peter K., and Rose McDermott. 2012. "A Neurobiological Approach to Foreign Policy Analysis: Identifying Individual Differences in Political Violence." *Foreign Policy Analysis* 8: 111-29.

Kosfeld , Michael, Markus Heinrichs, Paul J. Zak, Urs Fischbacher & Ernst Fehr. 2005. Oxytocin increases trust in human. *Nature* 435, 673-676

McDermott, R., Tingley, D., Cowden, J., Frazzetto, G. & Johnson, D. 2009. Monoamine Oxidase A gene (MAOA) predicts behavioral aggression following provocation. *Proceedings of the National Academy of Sciences*. 106 no. 7 2118-2123.

McDermott, R., C. Dawes, L. Prom-Wormley, L. Eaves, and P.K. Hatemi. forthcoming. "Maoa and Aggression: A Gene-Environment Interaction in Two Populations." Journal of Conflict Resolution

Meyer-Lindenberg, AM, Weinberger DR. 2006. Intermediate Phenotypes and genetic mechanisms of psychiatric disorders. *Nat Rev Neurosci* 7(10):818-27

Miles DR, Carey G. 1997. Genetic and environmental architecture of human aggression. J Pers Soc Psychol 72:207–217

Wersinger SR, Caldwell HK, Christiansen M, Young WS. 2007. Disruption of the vasopressin 1b receptor gene impairs the attack component of aggressive behavior in mice. Genes Brain Behav 6:653–660

Wood RM, Rilling JK, Sanfey AG, Bhagwagar Z, Rogers RD. 2006. Effects of tryptophan depletion on the performance of an iterated prisoner's dilemma game in healthy adults. Neuropsychopharmacology 31:1075–1084.

Part II: Implications of Aggressive Behavior

Chapter 4: Punishment and Reward

Anthony C. Lopez, Ph.D.
Assistant Professor of Political Science
Washington State University

Introduction

I will discuss the human psychology of reward and punishment in light of the evolutionary pressures that explain their function and operation, and in conjunction with the neurological and psychological evidence of the existence of these systems. Research in the area of reward and punishment is lopsided in two ways. First, most research focuses on the causes and consequences of *punishment* at the expense of an understanding of the causes and consequences of *reward*. This may be due to the fact that reward as a behavior-changing strategy is inherently unstable due to its vulnerability to exploitation, as I will discuss below. A second asymmetry in this research area is that reward and punishment are often examined exclusively in the context of *in-group behavior*. Comparatively little is understood about how reward and punishment function when the target of such (dis)incentives is explicitly a member of an *outgroup*.

Consequently, inferences to be drawn regarding the causes and consequences of punishment in a domestic context will be clearest. By contrast, inferences to be drawn regarding the causes and consequences of punishment – and especially reward – will be least clear and require the most caution when the context is conflict between political groups such as states.

State of Knowledge: Punishment and Reward Within Groups.

Punishment and reward cannot be understood outside the context of cooperation. Evolutionary biologists agree that any organism capable of cooperation must be able to defend against exploitation (Trivers, 1971; Axelrod & Hamilton, 1981). Two general strategies that a social organism can use to reverse and/or eliminate exploitation are the conferral of costs (punishment) and the conferral of benefits (reward). In short, the carrot or the stick. Cooperation is stable when defectors can be identified, excluded and/or punished, and when prospective cooperators can be identified, engaged, and rewarded through cooperative exchange. Much of this research takes for granted the larger context of such relationships occurring within one's group, and this is where we must begin.

Cooperation takes the form of a social contract in which one promises to pay a cost to receive a benefit, which is illustrated by the saying: "I'll scratch your back if you scratch mine"

(Dugatkin, 1997; Cosmides & Tooby, 2005). Punishment in the context of one-on one cooperation and group-level cooperation (i.e. "collective action") is reliably triggered when one perceives that another has received a benefit without paying the expected costs of participation. The literature has identified two general motivations for punishing free-riders in the context of collective action: 1) To sustain group norms that promote cooperation (Fehr & Gächter, 2002; Raihani & McAuliffe, 2012), or 2) To redress the welfare reduction experienced by participants relative to free-riders (Delton, Cosmides, Guemo, Robertson, & Tooby, 2012; Michael E Price, Cosmides, & Tooby, 2002). In the first explanation, individuals punish free-riders whenever group norms of prosociality and fairness are threatened, and the ultimate motivation is group welfare; in the second explanation, individuals are more likely to punish free riders when they themselves directly experience a reduction in welfare, and the ultimate motivation is individual welfare.

Although the above debate on causes is so far indeterminate, what is empirically the case is that individual-level behavioral and neurophysiological attributes explain much of the variation in willingness to punish free-riding, and when punishment is available, cooperation is often stable and free-riding deterred (Boyd & Richerson, 1992). Cooperation yields mutual gains that would not be possible without it. Unsurprisingly, therefore, individuals experience pleasure in cooperation and anger at being cheated. The very experience of cooperation in collective action itself, independent of task completion, is sufficient to trigger reward systems in the brain that encourage reciprocal exchange (Krill & Platek, 2012). By extension, those individuals who contribute the most to collective action are more likely to seek the punishment of free riders than those who have contributed less or not at all (Price et al., 2002; Price, 2005; Andreoni, Harbaugh, & Vesterlund, 2003). Punishing free-riders is more likely when individuals receive direct reputational benefits for these actions and when their behavior is monitored (Bateson, Nettle, & Roberts, 2006). High-testosterone males are more likely to punish unequal distributions of resources (Burnham, 2007), although testosterone also correlates with anti-social behaviors such as aggression and tendency to cheat (Zak et al., 2009). This generates the awkward empirical result that high-testosterone men are both more likely to punish those who are not generous toward them *and* less likely to be generous to begin with.

In human societies, the relevant question in response to undesired behavior is often not between punishment and reward, but between punishment and *reparation* (Petersen, Sell, Tooby, & Cosmides, 2012). Decisions regarding whether to punish an individual or repair the relationship depend on the value of the noncooperator to the group; decisions regarding the intensity of the response to the noncooperator depend upon the severity of the transgression. We seek to repair relationships with those we highly value, but punish others whose association is less valuable. For example, when a loved one or close friend commits a major transgression, this typically triggers an intense effort at reparation, not intense punishment. In contrast, when a stranger commits a major transgression, this typically triggers intense punishment, not an intense effort toward reparation. In both cases, the intensity of the reaction is high because the transgression is severe; but the *type* of reaction (punishment or reparation) depends on the association value of the target. The two key variables are

association value and severity of the transgression, which together determine the quality of a response to noncooperation.

These dynamics reflect an evolved psychology that is designed to expect the small-scale social interactions of ancestral environments (Kurzban & Neuberg, 2005; Sidanius & Kurzban, 2003). In these environments, social networks were dense and repeated. Reputation was critical and exclusion could be deadly. The ability to discriminate between potential cooperators and potential defectors was an intense selection pressure, and there is evidence that: humans possess specialized psychological systems for the purpose of detecting cheaters and free riders (Delton et al., 2012); humans are better than chance at predicting who is likely to cooperate based on the perception of facial features alone, such as facial width-to-height ratio, that tend correlate with underlying variations in testosterone (Verplaetse, Vanneste, & Braeckman, 2007; Yamagishi, Tanida, Mashima, Shimoma, & Kanazawa, 2003), and; free-riding reliably triggers neural circuits associated with negative emotions while the prospect of punishing free-riders activates reward centers in the brain, indicating emotional satisfaction as a consequence of their punishment (de Quervain et al., 2004).

Humans possess psychological systems designed to track potential cooperators/defectors and are armed with evolved intuitions and motivational systems regarding the proper response to cooperation and defection (punishment and reparation). Research on the role of reward in promoting cooperation is relatively lacking, most likely because research has failed to demonstrate a significant and reliable role of reward due to its vulnerability to exploitation (e.g. defectors accepting the reward and 'running'). To the extent there is a role for reward to play in promoting cooperation, it is emphasized *in conjunction* with punitive measures (Hilbe & Sigmund, 2010). In fact, economic experiments have revealed that when the opportunity to punish is absent, and only the opportunity to reward is available, levels of cooperation sink to levels that are actually lower than when neither punishment nor reward are available (Sefton, Shupp, & Walker, 2006). It is perhaps because of the vulnerability of reward strategies to exploitation that those who are especially generous are both more likely to give, *and to receive*, rewards for their generous behavior. In sum, reward may function less effectively as a behavior-*changing* strategy, but may function more effectively as a behavior-*sustaining* strategy.

Punishment is not doled out by cool and calculated rational behavior-modifying machines. As described above, the experience of exploitation in reaction to another's defection from cooperation or by the observation of free riders in the context of collective action reliably triggers anger, which primes behavioral strategies such as punishment, and in extreme circumstances, violence, against the defector or free rider. Emotional systems that enable anger have clear neural correlates (Archer, 1988; Harmon-Jones & Sigelman, 2001) and have been shown to be reliably triggered in the specific situations in which, through behavior or speech, another demonstrates that they value your welfare less than you believe they should (Sell, Tooby, & Cosmides, 2009). This often occurs when an individual's goal seeking behavior is thwarted by another (Fessler, 2010), or when one's ethical or moral codes are violated (Rozin, Lowery, Imada, & Haidt, 1999). The common theme here is that anger is reliably triggered when

another denies resources that you believe you deserve or are entitled to, whether the resource in question is reputation or material such as food. In this sense, anger is an evolved emotion system designed to alert the target of one's anger that they have undervalued your welfare and entitlements. When this happens at the group level (as discussed below) the result can be collective outrage, which functions as the group analogue of individual-level anger.

The above discussion indicates that:

- Individuals experience internal neurophysiological rewards for cooperating that bias individuals toward seeking cooperative relationships;
- Individuals possess innate systems designed to automatically scan for and distinguish cooperators from defectors based on visual/facial cues;
- The perception of free-riding triggers a motivation to punish especially by individuals who directly contribute in the collective action and care about their reputation (especially in monitored environments);
- The motivation to punish is facilitated by anger systems in the brain that are sensitive to threats of relative status; (further discussion below)
- Punishment-seeking will vary depending on individual-level attributes such as basal and circulating testosterone levels, as well as the presence of genetic markers in interaction with childhood trauma (McDermott, Tingley, Cowden, Frazetto, & Johnson, 2009);
- Severity of punishment will be a function of magnitude of the transgression, while a preference to repair the relationship will depend on the perceived association value of the targeted individual.

There is a literature that argues that patterns of punishment and reward are explained as a consequence of social reinforcement and observational learning (Seymour, Singer, & Dolan, 2007). That is: we know how and when to punish and reward others by learning from others and watching them. Laboratory studies reveal that animals can learn to fear actions and objects in their environment simply by viewing other animals display fear toward those actions or objects (Galef & Laland, 2005; Mineka & Cook, 1993). However, this literature may overstate the nature of brain plasticity by failing to recognize that some learning happens more easily than others, and some lessons are difficult to unlearn. Rats, for example, can easily learn to update their memory of food locations when those locations are experimentally altered; however, when locations of water are experimentally altered, they will persistently try to visit the previous location even in the face of evidence that water is no longer there. Evolutionarily, the location of food was more variable than water, and hence evolved reasoning mechanisms privilege the hypothesis that food location varies more than water location. Consequently, learning mechanisms are shaped by privileged hypotheses in the brain that structure the extent and quality of learning (Gallistel, 1990; Morgan, Rendell, Ehn, Hoppitt, & Laland, 2011). In humans, the *character* of costs imposed on others via punishment may vary culturally and be a function of social learning (e.g. what counts as a "cost"); however, as demonstrated above, punishment is reliably triggered by a certain set of environmental cues, mediated by specific emotions and neural substrates, and is directed at individuals with specific behavioral

attributes. These patterns are cross-cultural and supported by a range of behavioral and neuropsychological experiments.

Punishment and Reward Between Groups.

Punishment between individuals *and groups* takes the form of a withdrawal of benefits or the conferral of costs. If the former is chosen, the result is often to sever relationships with the outgroup; if the latter is chosen, the result is often violence, which is a form of punishment. Punishment between individuals *and groups* is mediated by emotions such as anger and hatred. The literature discussed above converges on the hypothesis that anger is an evolved system in the brain designed to resolve conflicts of interest in favor of the angry individual. In the case of violence, "killing one's antagonist is the ultimate conflict resolution technique" (Daly & Wilson, 1988). When violence is initiated for purely opportunistic reasons, it does not function as punishment. The goal of opportunistic violence is merely to take resources for personal gain (or prestige), not to seek redress for wrongdoing. This form of violence falls outside the scope of the current discussion. There is a longstanding set of findings from social psychology (Tajfel & Turner, 1986) and substantiated by animal and neuroscientific studies (Cikara, Botvinick, & Fiske, 2011; Mahajan et al., 2011) that the mere presence of outgroups is sufficient to trigger derogation of and competition with outgroups. This again, is not punishment; rather, these findings demonstrate that when inter-group punishment *does* occur, violations by outgroup members are likely to be punished more harshly than violations by ingroup members. Hormonal evidence reveals that coalitional context alone (i.e. is your opponent a member of your in-group or a member of an outgroup?) is sufficient to determine the extent of dominance striving in the face of competition between individuals (Flinn, Ponzi, & Muehlenbein, 2012). Thus, this one environmental indicator is sufficient to trigger a host of behavioral and physiological changes in the way individuals pursue competition and punishment.

The findings from the previous section would seem to lend favor to the view that punishment of free-riders is best explained as a function of individual-level welfare. However, this does not mean that concerns about group welfare do not play a role. In fact, such concerns may be especially powerful during inter-group conflict. During wartime, punishment is an especially effective strategy motivating participation, and the motivation does not seem to be personal gain, but group welfare (Mathew & Boyd, 2011). When at war, individuals are reliably more likely to punish non-contributors in their group than when they are not at war; furthermore, non-contributors and nonparticipants are more likely to express guilt and shame as a consequence of failing to participate in their group's war effort (Gneezy & Fessler, 2011; Puurtinen & Mappes, 2009). Interestingly, it is in the context of inter-group warfare that rewarding cooperation plays its most prominent role. In this context, reward less often takes the form of a transfer of material resources, and more often takes the form of the attribution of intangibles such as honor, valor, or higher social status (Glowacki & Wrangham, Forthcoming; Mathew & Boyd, 2011; Wrangham & Glowacki, 2012). In these contexts, perceptions of valor tend to be attributed to individuals who take great risks on behalf of the group, often independent of that individual's perceived strength and formidability (Keeley, 1996; Lopez,

Sznycer, & Petersen, In Prep). In short, between-group conflict promotes within-group cooperation, punishment of defectors, and guilt among defectors/non-participants. Importantly, however, it is likely that these effects are expected to be strongest when the in-group is *defensively responding* to threat rather than *initiating* violence against another group (Author Ph.D. Dissertation). It should come as no surprise therefore that the framing of warfare as either offensive or defensive is often hotly contested, since this cue alone affects the character of within-group cooperation, punishment, and reward.

As the above example shows, punishment in the context of group conflict cannot be understood absent the evolutionary logic of warfare between groups. Evolutionarily, warfare has most often taken the form of lethal raiding (Manson & Wrangham, 1991; Otterbein, 2004), and it has occurred in an ancestral environment that was "offense dominant," meaning it was easier to attack than to defend given the prehistoric state of weaponry combined with weak physical defenses (Gat, 2006). In this environment, there were sustained premiums for striking first via stealth and surprise, and for *not* being caught unaware and defenseless. This evolutionary pattern likely selected for a psychology biased toward threat sensitivity during between-group conflict in which "false alarms" are amplified and threats to group status are punished more severely than threats from ingroup members (Haselton & Buss, 2000; Navarrete, Kurzban, Fessler, & Kirkpatrick, 2004).

Given the fact that human psychology seems to reflect these ancestral pressures, it is not surprising that the most common motivation for warfare cross-culturally is revenge (Wrangham & Glowacki, 2012). In an offense dominant world, survival was a function of 1) the ability to strike first and, 2) the ability to credibly signal guaranteed retaliation. In this offense dominant world, spirals of violence are frequent, and deterrence through the promise of punishment functions to keep spirals from erupting (Walker & Bailey, 2012). In a world where offense is easy, the promise of punishment must be great – at least, sufficient to offset the benefits of a first-strike from one's adversary. These dynamics reveal the intimate connection between the evolutionary logic of warfare instantiated into human coalitional psychology and contemporary understandings of deterrence between states. Apropos, the heart of deterrence theory is the promise of unacceptable punishment (Brodie, 2007; Schelling, 1977). The "secure retaliatory force" that nuclear strategists argue is necessary for equilibrium in the nuclear age is nothing but a euphemism for "guaranteed vengeance," in which states promise a punishment that is greater than the benefits of striking first.

Although these dynamics are reflected in a human coalitional psychology that is sensitive to outgroup threats and that assigns disproportionate punishment for outgroup threats relative to ingroup threats, certain political and cultural environments may amplify this "offense dominant" mindset. One example is weak rule of law, in which there is uncertainty regarding the capability and legitimacy of third-party legal enforcement.

These situations often beget societies characterized by "culture of honor" traditions, in which, in the absence of capable and legitimate third-party enforcement, reputation for disproportionate retaliation/punishment becomes the most effective safeguard against

personal violence (Nisbett & Cohen, 1996). Consequently, we might expect those countries that are both isolated internationally, and characterized by "culture of honor" traditions domestically, to be especially concerned with the credibility of deterrents against rivals and to most actively seek the possession of nuclear weapons.

As mentioned above, inter-group punishment and violence are mediated by emotions such as anger and hatred. One recent historical analysis shows that political speeches by leaders that contained expressions of anger, contempt and disgust were reliably followed by acts of violence from that group (Matsumoto, Hwang, & Frank, 2012). Peaceful protests and acts of resistance were not preceded by speeches that contained cues to these emotions. Anger, contempt and disgust, however are likely each distinct emotions that prime different behavioral strategies. For example, a recent study distinguished between anger and hatred – the latter defined as represented by the idea of "stable negative characteristics in the out-group and the belief in the out-group's inability to undergo positive change" (Halperin, Russell, Dweck, & Gross, 2011, p. 276). This is noteworthy because hatred in this sense is an indication that *the relationship cannot be repaired*; namely, punishment, not reparation, is the only viable behavioral strategy. In a series of experiments in the context of Israeli-Palestinian violence, it was shown that inducing anger actually promoted compromise-seeking in negotiations; however, when anger was accompanied by hatred, subjects demonstrated decreased support for compromise (Halperin et al., 2011).

Concluding Remarks

The scientific literature on punishment and reward has mostly examined these behaviors and their neural correlates in the context of within-group interactions. However, some research in the context of between-group interactions are discussed, and inferences explored. These inferences are necessarily speculative, but additional research in these areas is forthcoming. What is clear is that humans possess specialized neural circuitry designed for navigating conflicts of interest, and that these operate in ways contingent upon individual and environmental attributes that are subject to variation. Punishment can promote cooperation, while the use of reward is most effective as a behavior-sustaining - not behavior-changing – device, especially in the context of defensive warfare. Punishment and reward should not be examined independently of the question of whether and how to *repair* relationships, which is often a direct function of an individual's value as an ally or group member.

References

Andreoni, J., Harbaugh, W., & Vesterlund, L. (2003). The Carrot or the Stick: Rewards, Punishments, and Cooperation. *American Economic Review*, *93*(3), 893–902. doi:10.1257/000282803322157142

Archer, J. (1988). *The behavioural biology of aggression*. Cambridge: Cambridge University Press. Retrieved from http://www.loc.gov/catdir/enhancements/fy0906/87021792-d.html

Axelrod, R., & Hamilton, W. D. (1981). The Evolution of Cooperation. *Science*, *211*(4489), 1390–1396.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, *2*(3), 412–414. doi:10.1098/rsbl.2006.0509

Boyd, R., & Richerson, P. J. (1992). Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups. *Ethology and Sociobiology*, *13*(3), 171–195.

Brodie, B. (2007). *Strategy in the Missile Age*. Rand Publishing.

Burnham, T. C. (2007). High-testosterone men reject low ultimatum game offers. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1623), 2327–2330. doi:10.1098/rspb.2007.0546

Cikara, M., Botvinick, M. M., & Fiske, S. T. (2011). Us Versus Them: Social Identity Shapes Neural Responses to Intergroup Competition and Harm. *Psychological Science*. doi:10.1177/0956797610397667

Cosmides, L., & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. *The Handbook of Evolutionary Psychology* (pp. 584–627). Hoboken, NJ: Wiley.

Daly, M., & Wilson, M. (1988). *Homicide*. New York: A. de Gruyter. de Quervain, D. J. F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U.,

Buck, A., & Fehr, E. (2004). The Neural Basis of Altruistic Punishment. *Science, New Series*, *305*(5688), 1254–1258.

Delton, A. W., Cosmides, L., Guemo, M., Robertson, T. E., & Tooby, J. (2012). The psychosemantics of free riding: dissecting the architecture of a moral concept. *Journal of personality and social psychology*, *102*(6), 1252–1270. doi:10.1037/a0027026

Dugatkin, L. A. (1997). *Cooperation among animals : an evolutionary perspective*. New York: Oxford University Press. Retrieved from http://www.loc.gov/catdir/enhancements/fy0637/96018864-d.html

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–140. doi:10.1038/415137a

Fessler, D. M. T. (2010). Madmen: An Evolutionary Perspective on Anger and Men's Violent Responses to Transgression. In M. Potegal, G. Stemmler, & C. Spielberger (Eds.), *International Handbook of Anger* (pp. 361–381). New York, NY: Springer New York. Retrieved from http://rd.springer.com/chapter/10.1007/978-0-387-89676-2_21

Flinn, M. V., Ponzi, D., & Muehlenbein, M. P. (2012). Hormonal Mechanisms for Regulation of Aggression in Human Coalitions. *Human Nature*. doi:10.1007/s12110-012-9135-y9

Galef, B., & Laland, K. (2005). Social Learning in Animals: Empirical Studies and Theoretical Models. *BioScience*, *55*(6), 489–499.

Gallistel, C. R. (1990). *The organization of learning*. Cambridge, Mass.: MIT Press.

Gat, A. (2006). *War in Human Civlization*. Oxford: Oxford University Press. Retrieved from http://www.loc.gov/catdir/toc/ecip0614/2006017223.html

Glowacki, L., & Wrangham, R. (Forthcoming). The role of rewards in motivating participation in simple warfare: A test of the cultural rewards war-risk hypothesis. *Proceedings of the Royal Society B: Biological Sciences*.

Gneezy, A., & Fessler, D. M. T. (2011). Conflict, sticks and carrots: war increases prosocial punishments and rewards. *Proceedings of the Royal Society B:Biological Sciences*. doi:10.1098/rspb.2011.0805

Halperin, E., Russell, A. G., Dweck, C. S., & Gross, J. J. (2011). Anger, Hatred, and the Quest for Peace: Anger Can Be Constructive in the Absence of Hatred. *Journal of Conflict Resolution*, *55*(2), 274–291. doi:10.1177/0022002710383670

Harmon-Jones, E., & Sigelman, J. (2001). State anger and prefrontal brain activity: evidence that insult-related relative left-prefrontal activation is associated with experienced anger and aggression. *Journal of personality and social psychology*, *80*(5), 797–803.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: a new perspective on biases in cross-sex mind reading. *Journal of personality and social psychology*, *78*(1), 81–91.

Hilbe, C., & Sigmund, K. (2010). Incentives and opportunism: from the carrot to the stick. *Proceedings of the Royal Society B: Biological Sciences*. doi:10.1098/rspb.2010.0065

Keeley, L. H. (1996). *War before Civilization: The Myth of the Peaceful Savage*. New York: Oxford University Press. Retrieved from http://www.loc.gov/catdir/enhancements/fy0638/94008998-d.html

Krill, A. L., & Platek, S. M. (2012). Working Together May Be Better: Activation of Reward Centers during a Cooperative Maze Task. *PLoS ONE*, *7*(2), e30613. doi:10.1371/journal.pone.0030613

Kurzban, R., & Neuberg, S. (2005). Managing Ingroup and Outgroup Relations. *The Handbook of Evolutionary Psychology* (pp. 653–675). Hoboken: John Wiley & Sons.

Lopez, A. C., Sznycer, D., & Petersen, M. B. (In Prep). The measure of valor: Risk and formidability in war. Mahajan, N., Martinez, M. A., Gutierrez, N. L., Diesendruck, G., Banaji, M. R., & Santos, L. R. (2011). The evolution of intergroup bias: perceptions and attitudes in rhesus macaques. *Journal of personality and social psychology*, *100*(3), 387–405. doi:10.1037/a0022459

Manson, J. H., & Wrangham, R. W. (1991). Intergroup Aggression in Chimpanzees and Humans. *Current Anthropology*, *32*(4), 369–390.

Mathew, S., & Boyd, R. (2011). Punishment sustains large-scale cooperation in prestate warfare. *Proceedings of the National Academy of Sciences*, *108*(28), 11375– 11380. doi:10.1073/pnas.1105604108

Matsumoto, D., Hwang, H. C., & Frank, M. G. (2012). Emotions expressed in speeches by leaders of ideologically motivated groups predict aggression. *Behavioral Sciences of Terrorism and Political Aggression*, *0*(0), 1–18. doi:10.1080/19434472.2012.716449

McDermott, R., Tingley, D., Cowden, J., Frazetto, G., & Johnson, D. D. P. (2009). Monoamine Oxidase A Gene (MAOA) Predicts Behavioral Aggression Following Provocation. *Proceedings of the National Academy of Sciences*, *106*(7), 2118–2123.

Mineka, S., & Cook, M. (1993). Mechanisms involved in the observational conditioning of fear. *Journal of experimental psychology. General*, *122*(1), 23–38.

Morgan, T. J. H., Rendell, L. E., Ehn, M., Hoppitt, W., & Laland, K. N. (2011). The evolutionary basis of human social learning. *Proceedings of the Royal Society B: Biological Sciences*. doi:10.1098/rspb.2011.1172

Navarrete, C. D., Kurzban, R., Fessler, D. M. T., & Kirkpatrick, L. A. (2004). Anxiety and Intergroup Bias: Terror Management or Coalitional Psychology? *Group Processes & Intergroup Relations*, *7*(4), 370–397. doi:10.1177/1368430204046144

Nisbett, R. E., & Cohen, D. (1996). *Culture of honor : the psychology of violence in the South*. Boulder, Colo.: Westview Press. Retrieved from http://www.loc.gov/catdir/enhancements/fy0832/96166629-b.html

Otterbein, K. F. (2004). *How War Began* (1st ed.). TAMU Press.

Petersen, M. B., Sell, A., Tooby, J., & Cosmides, L. (2012). To punish or repair? Evolutionary psychology and lay intuitions about modern criminal justice. *Evolution and Human Behavior*. doi:10.1016/j.evolhumbehav.2012.05.003

Price, M. E. (2005). Punitive sentiment among the Shuar and in industrialized societies: cross-cultural similarities. *Evolution and Human Behavior*, *26*(3), 279–287. doi:10.1016/J.Evolhumbehav.2004.08.009

Price, Michael E, Cosmides, L., & Tooby, J. (2002). Punitive Sentiment as an Anti-Free Rider Psychological Device. *Evolution and Human Behavior*, *23*(3), 203–231.

Puurtinen, M., & Mappes, T. (2009). Between-group competition and human cooperation. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1655), 355 –360. doi:10.1098/rspb.2008.1060

Raihani, N. J., & McAuliffe, K. (2012). Human punishment is motivated by inequity aversion, not a desire for reciprocity. *Biology Letters*, *8*(5), 802–804. doi:10.1098/rsbl.2012.0470

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: a mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of personality and social psychology*, *76*(4), 574–586.

Schelling, T. C. (1977). *Arms and Influence.:* Greenwood Press Reprint.

Sefton, M., Shupp, R., & Walker, J. M. (2006). The Effect of Rewards and Sanctions in Provision of Public Goods. *SSRN eLibrary*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=932683

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the Logic of Human Anger. *Proceedings of the National Academy of Sciences*, *106*(35), 15073–15078.

Seymour, B., Singer, T., & Dolan, R. (2007). The neurobiology of punishment. *Nature reviews. Neuroscience*, *8*(4), 300–311. doi:10.1038/nrn2119

Sidanius, J., & Kurzban, R. (2003). Evolutionary Approaches to Political Psychology. *Oxford Handbook of Political Psychology*. New York: Oxford University Press.

Tajfel, H., & Turner, J. C. (1986). The Social Identity Theory of Intergroup Behavior. *Psychology of Intergroup Relations*. Chicago: Nelson Hall.

Trivers, R. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, *46*(1), 35–57.
Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: the sequel. A kernel of truth in predictive cheating detection. *Evolution and Human Behavior*, *28*(4), 260–271. doi:10.1016/j.evolhumbehav.2007.04.006

Walker, R. S., & Bailey, D. H. (2012). Body counts in lowland South American violence. *Evolution and Human Behavior*. doi:10.1016/j.evolhumbehav.2012.08.003

Wrangham, R. W., & Glowacki, L. (2012). Intergroup Aggression in Chimpanzees and War in Nomadic Hunter-Gatherers : Evaluating the Chimpanzee Model. *Human Nature (Hawthorne, N.Y.)*. doi:10.1007/s12110-012-9132-1

Yamagishi, T., Tanida, S., Mashima, R., Shimoma, E., & Kanazawa, S. (2003). You Can Judge A Book by its Cover: Evidence That Cheaters May Look Different From Cooperators. *Evolution and Human Behavior*, *24*(4), 290–301.

Zak, P. J., Kurzban, R., Ahmadi, S., Swerdloff, R. S., Park, J., Efremidze, L., Redwine, K., et al. (2009). Testosterone Administration Decreases Generosity in the Ultimatum Game. *PLoS ONE*, *4*(12), e8330.

Chapter 5: Threat Perception and Deterrence

Rose McDermott, Brown University, Rose_McDermott@Brown.edu
Peter K. Hatemi, Pennsylvania State University, phatemi@gmail.com

Some human needs appear universal, or nearly so. Among these are desires for status, reputation, attention, appreciation, prestige, and even glory. And some human emotions, including fear and anger, seem nearly ubiquitous as well. Yet, in the midst of such common urges lie vast individual differences in the baseline propensity for experiencing such phenomena as well as the particular environmental cues and triggers which might spark or diminish them. As we examine the effect of these factors on threat perception, and responses

to both challenges and opportunities in the physical and social environments that surround us, we provide not so much definitive answers as novel questions which offer a different perspective providing unique insight into the role of emotion in decision making around violence.

The recognition of threat is infused with an array of emotions which can conflict and collide in making decisions about what to do in response to it. Because of this, one of the best ways to prevent having to make potentially costly decisions under conditions of high risk and time urgency is to do whatever can be done to prevent the threat in the first place. This is where the notion of deterrence can come into play. And the underpinnings of deterrence can be found in basic human psychological architecture. Because challenges or predation and out-group threat have faced humans for millennia, analyzing the notion of deterrence from the perspective of evolutionary models may prove helpful; examining the genetic and biological mechanisms which precipitate our recognition and response to threat can inform our understanding of how to create more accurate signals and more effective responses. Indeed, much of the insight of this perspective emerges from the fundamental logic of revenge that permeates the existence of first strike coalition-dwelling creatures such as ourselves, which humans have been for millennia.

Long before nuclear weapons appeared, deterrence as a concept was baked into our nature, and part of our psychology. Indeed, such basic strategies can be found in our non-human, primate ancestors. This does not mean that humans are incapable of rational cost-benefit analysis, or of overcoming their visceral instinctual responses. Quite the contrary, to deter is rational. It only means that humans still retain those automatic responses, and experience those thoughts and feelings as powerful, real information which can influence the stimuli they pay attention to and the kinds of responses they want to make in reaction to threat or attack. Thus, in addition to whatever rational strategic thinking in which we engage when we confront the strategic challenges of the modern world, we still often process the world using our inherent biology and ancestrally relevant cues. These cues are inherently rational. Fighting has always been both costly and risky, but also enticing and endogenously reinforcing to some. All rational actors have an interest in settling things with threats but without the use of violence. Thus, deterrence is neither new nor is it an invention of nuclear strategists. The fact that deterrence has been linked with nuclear weapons should not distract from the reality that deterrence is a recurrent problem that any coalition must deal with in hostile or competitive environments, such as when confronting threatening out-groups. In this regard, and more generally, there is an intimate link between deterrence, reputation, and credibility.

It follows that both bluffing and detection of deterrence constitutes part of our natures as well. This would suggest that evolution must have involved an arms race between these capabilities in designing our neurocomputational mental architecture. The role of such a competition helps illuminate the role of overconfidence in male coalitionary psychology (Johnson, 2006). Wrangham (1999a) and Trivers (2011) argue that it is easier to bluff others if we deceive ourselves because there will be less behavioral leakage. If we really believe we will win a fight, it is easier to convince others to join, since most people will want to join a winning

campaign. The greater ease in labor recruitment for combat in and of itself increases the likelihood of victory because, at least in many fights, and especially those prior to the onset of mechanized weaponry, size, along with surprise, would have constituted a definitive factor in the probability of success. And if a strong enough bluff can make the other side back down prior to actual combat, such action only enhances the success of the strategy at very low cost. In this way, overconfidence can enhance the number of people who want to join a coalition, making it more likely that that group would be able to win a fight, but also make it less likely that such a fight will actually take place.

However, no strategy is entirely without costs and individuals often have difficulty seeing or making the necessary value trade-offs, which exist in realms that go far beyond the guns versus money divide, in seeking national security in response to threat. For example, actions in support of one ally, like Israel, might alienate other allies, or further alienate those already predisposed against American interests. Other related self-serving psychological biases such as wishful thinking, also capture the basic human desire to avoid psychological pain associated with loss and death. Such biases may also prevent the recognition of threat or reduce the ability to formulate an appropriate and timely response to it. A classic example of this dynamic comes from Neville Chamberlain's attempt to placate Hitler at Munich rather than build the defenses for the imminent war which would soon engulf all of Europe; had Chamberlain proved willing to build such defenses and fight sooner, Germany would not been in as strong a position to seize the initiative in Operation Barbarossa in September, 1939.

In addition, particular motivational biases can lead one actor to be sure that his deterrent threat is credible while simultaneously leading the adversary to misperceive or downgrade that same threat. This can happen when a leader threatens a particular response if an offender crosses a certain line and then fails to deliver the promised consequence. In a similar manner, the need to address multiple audiences can also confuse the meaning of particular threats or reassurances. For example, a message designed to reassure a domestic audience may simultaneously threaten an opponent unnecessarily. American presidential campaigns are often filled with bluster against other states, for example, that only rarely have a credible basis in actual policy planning. However, a nation may dismiss a serious threat precisely because the challenge is dismissed as intended solely for a domestic audience and does not constitute a serious threat. This may have been part of the reason that the Americans were unprepared for the Chinese counterattack over the Yalu River during the Korean War. Such lack of proper calibration may leave the defender dangerously vulnerable to an attack for which they may have otherwise had adequate time to mount an adequate defense.

The Psychology of First Strike, Coalitionary Humans and Maximum Response

Azar Gat (2006) discusses deterrence briefly in his book, *War and Human Civilization*. He argues that deterrence is that which keeps spirals of violence in check. In the realm of interpersonal violence, Gat argues that weak physical defenses combined with sophisticated tool making which generated increasingly powerful shock and fire weapons meant that ancestral humans were basically in an "offense dominant" position. Because of this, he says

that ancestral humans were "quintessential first-strike creatures." If this dynamic asymmetry between weak physique and strong weaponry persisted long enough in ancestral environments, psychological incentives would have come to favor first-strike capability over time. In other words, the coalitional psychology of combat would have favored specializations for surprise attack, lethal raids, and so on. In addition, individuals would have invested a great deal in developing strong reputations for disproportionate response to attack as a deterrent against revenge attacks (vengeance). One of the best examples comes from Nixon, who used such logic to undergird his Madman Theory, in seeking an end to American military involvement in Vietnam. One day, walking along the beach in California, shrouded in fog, he told his Chief of Staff Bob Haldeman (1978, 83), "I call it the Madman theory, Bob. I want the North Vietnamese to believe I've reached the point where I might do anything to stop the war. We'll just slip the word to them that, 'for God's sake, you know Nixon is obsessed about Communism. We can't restrain him when he's angry–and he has his hand on the nuclear button'–and Ho Chi Minh himself will be in Paris in two days begging for peace." While this strategy did not appear to work for Nixon, he believed that it would. Like Schelling's (1960) threat that leaves something to chance, or his notion of the rationality of irrationality, Nixon believed that creating a reputation for disproportionate response would advantage his play against an adversary by encouraging them to back down in the face of threat.

Dovetailing on the above, and independent of his central contention that violence has decreased over time in human history, Steve Pinker (2011) in his recent book "*Better angels of our nature*" notes that the "most commonly cited motive for warfare is vengeance, which serves as a crude deterrent to potential enemies by raising the anticipated long-term costs of an attack." While much of this work remains controversial and has yet to be supported by alternative lines of evidence, the proposition that violence has played a central role in human history poses a critical insight into the evolutionary basis for the establishment of a biological underpinning for aggression and violence in the face of threat. As we know from Wrangham's (1999b) work with chimpanzees and the critical importance of the 3 to 1 imbalance of power ratio, the elimination of even a few adult males from the group can be a critical blow in a world of small-scale coalitional hostility. Wrangham shows that chimpanzees who confront each other at territorial borders will try to run away from each other unless one group has a three to one numerical advantage over the other; when this imbalance of power occurs, the stronger group will try to overpower and kill the smaller group. Where vengeance is certain to be provoked by an attack, deterrence kicks in when the initiators cannot be absolutely sure that they'll be successful. As Pinker says, "That is why they sometimes massacre every last member of a village they raid: they anticipate that any survivors would seek revenge for their slain kinsmen." This, in effect, is the nuclear option - ancestral style. Attackers seek to eliminate the enemy's retaliatory ability. One side destroys the other out of fear of reprisal; over time, this kind of behavior would have been strong incentives since more conciliatory strategies would have become rendered such groups extinct. In other words, in the offense-dominant world of our ancestors, where vengeance is a predictable component of the human repertoire, your best bet is maximum response. Although humans now must operate in large scale industrialized societies and cultures, and consider rational responses to incipient threats, the instinctual desires for vengeance and maximum response in reaction to provocation and threat may still

remain, even if it can be over-ridden by more rational calculations in most instances. However, there are certainly those who would argue that the Bush administration attack on Iraq in 2003 was driven more by vengeance than rational calculation. This vengeance may have received support on the part of an American public who, by and large, did not understand or did not care that Iraq had nothing to do with the attacks on 9/11, but the desire for revenge for other reasons may have motivated at least some members of the administration. In the end, it is important to note that American military reaction in response to this event actually served the best interests of Al Qaeda: America nearly bankrupted itself conducting two wars and establishing many new bureaucratic agencies (a strategy Reagan had used to undermine the Soviet state in the 1980s); this action also provided relatively easy targets in Iraq and Afghanistan against which to hone their skills, train their soldiers, and test their strategy; it increased their status throughout the Arab world, enhancing their ability to recruit from the ranks of countries where they had no previous ties.

A strategy of maximum response in reaction to threat begins to look a lot like the logic of deterrence as it became crystalized during the nuclear age. The highly sharpened costs of nuclear warfare only clarified what for millennia had been deeply engrained principles of coalitional violence. Ancestral coalitional environments appear to have been dominated by cultures of perpetual offense dominance as Wrangham's imbalance of power in chimps demonstrates. Coalitionary violence provides the psychological, if not military basis, for what Brodie (1962) and others called "secure retaliatory force." He and his colleagues argued that what was essential for deterrence was nothing short of a euphemism for "guaranteed vengeance," or what we have just been describing in terms of maximum response in order to develop an adequate reputation for resolve, and thus prevent attack in the first place. In this sense, deterrence is achieved when a country believes that the cost of "guaranteed vengeance" from its enemy is too great to instigate an attack from the outset. This, in turn, is not too far from Schelling's notion that the underlying force behind bargaining and coercive diplomacy is the "power to hurt." Of course modern deterrence theory is more complicated than this, but the basic psychological dynamics underling debates over counter-force versus counter-value targeting, and the nature of the stability-instability paradox, emanate from the same desire for maximum response in the face of threat.

Pinker also makes another interesting point about the contribution of culture to the basic dynamics underlying deterrence that is worth noting. He observes that the vengeance/reputation/deterrence dynamic is especially powerful in societies where the rule of law is weak, anarchy prevails, and cultures of honor are embedded. In this regard, it would be interesting to examine whether any correlations exist between domestic factors and foreign policy in such nations. Is there a relationship between: 1) those countries striving hardest to acquire nuclear weapons (or other forms of status in the form of weaponry); and 2) various political and cultural cues, such as weak rule of law and cultures of honor. In other words, the combination of nuclear weapons, weak states and cultures of honor suggest the possibility that the kind of deterrence that proves most effective for some cultures might be different than the kind that works best for other cultures. Moreover, it may be the case that particular kinds of institutions and cultures are more likely to seek opportunities for proliferation precisely

because of these structural incentives. If this is the case, then interventions designed to strengthen such institutions may prove an efficient and effective means toward reducing the likelihood of proliferation of weapons of mass destruction.

Leadership and Group Dynamics

There are two specific domains in which applications of these notions of threat perception and deterrence hold concrete relevance and potential application. The first has to do with leadership and the second with group dynamics and the possibility of predicting or reducing incidences of violence.

With regard to leadership, the more we are able to learn about individual behavior, and the sources of individual variance in its origin, the higher probability we will be able to predict how a given individual will act under duress. When combined with enhanced knowledge of group behavior, as described further below, and information about the resource constraints under which any given leader is acting, the more possible it becomes to monitor potential breaking points. If we can create a more neurobiologically informed understanding of individual dispositions and personal psychology, it may become possible to locate those triggers that cue a particular individual to respond in a hostile as opposed to conciliatory manner in the face of threat, and structure interventions which might provide a firewall break between provocation and response in such as manner as to enhance the possibility for more constructive interactions between leaders and states.

Because, even in a world of states, there is still someone who has to decide to push the button; understanding better how and why someone might come to do that would behoove us all. After all, why do people want to have nuclear weapons? For many states, there should be a great rational interest in not doing so because it costs lots of money and takes time and energy to develop such programs. In addition, their existence generates negative responses from other states. And yet leaders and their population often still appear to want to have them. Other states that are strong enough may be able to constrain them through the distribution of power, and such efforts may be able to shift behavior in the short run, but it remains unlikely that institutional and structural constraints will be able to change the underlying motivation and desire for such weapons, which are likely rooted not only in those universal desires for status and reputation but also in the drive to be able to exert maximum response under conditions of threat.

Most studies of political leadership have tended to use historical case study analyses which relied on psychoanalytic theory, or other forms of analysis of leaders from a distance (Hermann, 1980; Post, 1991). This work often involved archival or interview work and attempted to focus on personality (Lasswell, 1930; George & George, 1956), For example, Lasswell (1930) argued that leaders project their personal struggles onto the larger political world in which they operate. Barber (1972) provides perhaps one of the better models of the relationship between personality and presidential leadership, focusing on leaders' character, worldview and style. Greenstein (1967) argued that leaders can have an impact on their

environment to the extent that the environment can be restructured given a leader's particular strengths and weaknesses; certain situations are more likely to provide a match with particular leaders' personalities. These studies proved intriguing and some were more beneficial than others, but all are researcher dependent, not replicable, and only offer post-hoc prediction. Other studies have focused on surveys. Etheredge (1978) administered personality batteries to 36 State Department officials and then correlated their responses with their tendency to use force in 49 crises in American Foreign Policy. He was able to predict their responses to the crises with greater than 75% accuracy based on the personality inventories. Those who advocated the greatest use of military force in response to foreign policy challenges were those most likely to show high dominance displays toward their underlings at work, revealing a systematic patterns of response styles across personal and professional domains. Other models attempted to understand leaders not by focusing on leaders but by identifying the basic psychological mechanisms undergirding leadership dynamics using experimental methods (Lewin, Lippett &White, 1939). The difficulty with all of these approaches lies in the inability to model a single individual who makes the decision. That is, it is impossible to survey Pol Pot, or Ahmadinejad. Furthermore, while all actions can be identified post-hoc, understanding the source of the motivations remains hidden.

An alternative line of reasoning which has focused on evolutionary and neurobiological perspectives has served to enlighten aspects of leadership which remain hidden from the perspective offered by traditional models. A genetically informed evolutionary view can help explain the source of those motivations. A great deal of attention has been paid to the question of leadership being either born or made. Certainly it is both, but neurobiological methods can provide greater specificity about the interaction of genes and environment in creating leaders. A series of behavioral genetic studies has found that genetic influences on leadership account for roughly 30-44% of the variance (Arvey et al 2006, 2007; Chaturvedi et al (2011). van Vugt (2006) suggests that leaders also emerge in the face of substantial threats or opportunities, since leaders are typically the people who move first in such situations. Such leadership provides the benefits of coordinated action which prove most helpful in times of stress of threat. In other work using real world leader samples, Carnevale et al. (2011) find that leaders who scored higher on the need for cognition, meaning they seek out and enjoy cognitive effort, performed better on tasks related to decision making competence. These leaders outperformed controls, suggesting that leadership, if only in this example, actually does reflect some aspect of increased skill or ability, at least in some domains. To further examine the novel hypotheses an evolutionary perspective can bring to leadership studies, Van Vugt et al. (2008) argue that leadership and followership evolved in the ancestral environment to help overcome the repeated challenges associated with social coordination problems, including the need for collective action. These repeated problems included the need for group movement, intragroup cohesion and successful intergroup competition. They note the inherent tension between the need for effective coordinated action, as potentiated by leadership, and the possibility that such action allows for the exploitation of followers, introducing the enduring ambivalence between leaders and followers. Additional work drew upon Lewin, Lippett and White's (1939) paradigm described above, showing that members are more likely to leave groups with autocratic as opposed to democratic or laissez-faire leadership styles. Importantly,

van Vugt et al. (2004) found that such effects held regardless of the personal resources members derived from leaders, indicating that their objections to an autocratic leadership style resulted from procedural as opposed to distributive reasons. This finding runs contrary to arguments made by Bueno de Mesquita et al. (2003) and others which explicitly focus on how leaders stay in power through their ability to differentially distribute resources to their winning coalition members. By contrast, in the van Vugt studies, individuals preferred democratic leaders who had a legitimate power base. This preference appeared much stronger when group identity remained high, in which case either instrumental or relational leaders proved equally efficient at garnering contributions from followers. However, when group identity was low, instrumental leaders were more effective at obtaining such benefits from members (van Vugt & De Cremer, 1999), possibly explaining those conditions under which the Bueno de Mesquita et al. (2003) model holds true.

Many of these leadership processes appear potentiated by precisely the biological factors and precipitants we endorse examining in leadership studies. Luizza et al., (2011) relied on the use of eye tracking technology to examine how the gaze of political leaders affects the gaze of in-group and out-group followers. The authors hypothesized these relationships based on primate literature which suggested that the automatic tendency to follow the gaze of other group members can be affected by relative social status. In this study, researchers examined the directional gaze of right wing Italian leader Silvio Berlusconi. They found that in-group members followed his gaze whereas out-group members tended not to look where he was looking. In this way, a leader's gaze proved predictive of seemingly reflexive shifts in attention; this bias could either result from increased affiliation with in-group leaders, or simply reflect shared differences in attentional bias between leaders and followers of the same political persuasion.

Other techniques have been harnessed to identify the specific neurological, genomic, or hormonal systems that account for this variance. Baltazard et al. (2011) utilized electroencephalography (EEG) to differentiate transformational leaders from non-transformational leaders on the basis of this additional neurobiological tool. In this way, specific leadership traits were shown to be related to different levels of electrical activity in different parts of the brain. Additional work employing another technique involving hormonal assays to explore the endocrinology of leadership has also brought new light to bear on the neurobiology of leadership. For example, work conducted by Robert Josephs and colleagues finds a relationship between testosterone and social status, demonstrating that in high status positions, high testosterone individuals do well regardless of task content, whereas they perform poorly on both spatial and verbal tasks when placed in a low status situation (Newman et al., 2005). In addition, the effect of testosterone also appears mediated by the role of cortisol, a stress hormone, as well. Mehta & Josephs (2010) find a relationship between testosterone and dominance, but only in those individuals with low cortisol. When cortisol was high, the relationship between testosterone and dominance disappeared, or reversed. This suggests a reason for the lack of relationship between leadership and dominance reported in the Van Vugt review noted above; it is entirely possible that individuals in Van Vugt's review had high cortisol, a plausible conclusion if many subjects emerged from student samples who

found the experience of leadership stressful. In the Mehta & Josephs (2010) work, neuroendocrine effects appeared particularly pronounced under conditions of social treat or social defeat. This work offers truly profound evidence in support of an evolutionary hypothesis establishing a foundation for leadership by delineating the hormonal link between the reproductive (testosterone) and stress (cortisol) pathways in regulating dominance displays and behavior. Their work suggests that because testosterone potentiates status seeking in social hierarchies, only when stress and threat are low will high testosterone lead to higher status; when stress is high, high testosterone may instead be associated with lower status. This work demonstrates the kinds of novel insights into leader decision making and behavior that becomes possible using the tools and insights garnered from a neurobiological perspective. Josephs et al. (2006) suggest that it is precisely this mismatch between biological reality as embodied in testosterone and social status that can lead to dysfunction, discomfort and disease. When low testosterone individuals are placed in high status positions, they display greater emotional arousal, including higher heart rate, and poorer cognitive performance, just as occurs when high testosterone people are placed in low status positions. This suggests that particular individuals may be both more predisposed and more able to assume leadership roles. But such a tendency may only manifest under particular environmental circumstances involving threat or opportunity for members of an in-group which holds high salience and meaning for participants. Under such circumstances, an incipient leader can then draw on relational skills and abilities to leverage social identity to overcome collective action challenges. Finally, Doug Madsen, in a series of prescient experiments in the 1980's provided a remarkable empirical demonstration of the use of whole blood serotonin to predict power seeking drives, defined as striving for social dominance, among individuals. Madsen's work related this biochemical marker to several behavioral patterns, including aggressiveness, competitiveness and distrust. This work constituted the first, and so far only, clear documentation of a biochemical marker to discern differences among individuals in a critical area directly related to leadership drive. Such novel theoretical and methodological approaches can further deepen our understanding of the neurobiological processes undergirding such phenomena and help illuminate the basis of important characteristics that potentiate good leadership or precipitate poor leadership. As should be evident, existing studies of leadership have no comprehensive model of individual variance. Employing a neurobiological perspective could help provide a model which combines both biological and environmental forces into a more cohesive model of how individual dispositions might inform political choice. In this way, the existing variety of theoretical perspective can be enhanced by incorporating a neurobiological approach though obtaining DNA or saliva samples on subjects. Whereas before leadership studies remained largely idiosyncratic and anecdotal, new methods allows scholars the ability to map physiology into psychobiography in an integrated fashion which can provide a more holistic understanding and representation not only of individual leaders, but the nature of leadership itself.

By exploring the foundations of human psychological and biologically informed notions of threat and deterrence from a neurobiologically informed perspective, we can begin to leverage our own biology in service of our very survival through a recognition of those environmental cues and triggers which both instigate and extinguish our desires for aggression and cooperation.

References

Arvey, Richard, Maria Rotundo, Wendy Johnson , Zhen Zhang, Matt McGue. 2006. The determinants of leadership role occupancy: Genetic and personality factors. The Leadership Quarterly 17: 1– 20

Arvey, Richard, Zhang, Zhen, Avolio, Bruce & Krueger, Robert. 2007. Developmental and Genetic Determinants of Leadership Role Occupancy Among Women *Journal of Applied Psychology* 92 (3):693–706

Baltazard , Pierre. Waldman, David, Thatcher, Robert & Hannah, Sean. 2011. Differentiating transformational and non-transformational leaders on the basis of neurological imaging *The Leadership Quarterly ISSN 1048-9843, 10.1016/j.leaqua.2011.08.002.*

Barber, James David. 1972. *The Presidential Character: Predicting Performance in the White House.* Englewood Cliffs, NJ: Prentice-Hall.

Brodie, Bernard. 1962. Defense Policy and the Possibility of Total War. *Daedulus*, 91 (4), 733-

Carnevale, Jessica, Yoel Inbar, Jennifer S. Lerner. 2011. Individual differences in need for cognition and decision-making competence among leaders Personality and Individual Differences 51 (3): 274-278

Bueno de Mesquita, Bruce, Smith, Alastair, Siverson, Randolph & Morrow, James. 2003. *The Logic of Political Survival.* Cambridge, MA: MIT Press.

Etheredge, Llyod. 1978. Personality Effects on American Foreign Policy, 1989-1968. *American Political Science Review* 72(2): 434-51.

Gat, Azar. 2008. *War and Human Civilization.* Oxford: Oxford University Press.

George, Alexander & Juliette. 1956. *Woodrow Wilson and Colonel House: A Personality Study.* New York: Dover.

Greenstein, Fred. 1967. The Impact of Personality and Politics. *American Political Science Review* 61: 629-41.

Haldeman, H. R. with Joseph DiMona. 1978. *The Ends of Power*. New York: Times Books.

Hermann, Margaret. 1980. Explaining Foreign Policy Behavior using the personal characteristics of political leaders. *International Studies Quarterly* 24: 7-46

Josephs, Robert A.; Sellers, Jennifer Guinn; Newman, Matthew L.; Mehta, Pranjal H. 2006. The mismatch effect: When testosterone and status are at odds. *Journal of Personality and Social Psychology*, 90(6): 999-1013.

Lasswell, Harold. 1930. *Psychopathology and Politics.* Chicago: University of Chicago Press.

Lewin, K, R. Lippett & R. White. 1939. Patterns of aggressive behavior in experimentally created 'social climates." *Journal of Social Psychology* 10: 271-99.

Liuzza MT, Cazzato V, Vecchione M, Crostella F, Caprara GV, et al. 2011. Follow My Eyes: The Gaze of Politicians Reflexively Captures the Gaze of Ingroup Voters. *PLoS ONE* 6(9): e25117.

Madsen, D. 1986. ''Power Seekers are Different: Further Biochemical Evidence.'' American Political Science Review 80: 261–69.

Mehta, Pranjal H.; Josephs, Robert A. . 2010. Testosterone and cortisol jointly regulate dominance: Evidence for a dual-hormone hypothesis. *Hormones and Behavior*, 58(5): 898-906.

Post, Jerrold. 1991. Saddam Hussein of Iraq: A political psychology profile. *Political Psychology* 12: 279-89.

Pinker, Steven. 2011. *The Better Angels of our Nature: Why Violence has Declined.* New York: Viking.

Schelling, Thomas. 1960. *The Strategy of Conflict.* Cambridge: Harvard University Press.

Trivers, Robert. 2011. The Evolution and Psychology of Self Deception. *Behavior and Brain Sciences* 34 (1): 1-16.

Van Vugt, M., Jepson, S., Hart, C. and De Cremer, D. . 2004. Autocratic leadership in social

dilemmas: A threat to group stability. *Journal of Experimental Social Psychology*, 40, (1), 1-13.
Van Vugt, Mark; De Cremer, David . 1999. Leadership in social dilemmas: The effects of group identification on collective actions to provide public goods. *Journal of Personality and Social Psychology,* 76(4): 587-599.
Van Vugt, Mark; Spisak, Brian R. 2008. Sex differences in the emergence of leadership during competitions within and between groups. *Psychological Science* 19(9):854-858.
Van Vugt, Mark; Hogan, Robert; Kaiser, Robert B. 2008. Leadership, followership, and evolution: Some lessons from the past. *American Psychologist*, 63(3): 182-196.
Van Vugt, Mark . 2006. Evolutionary Origins of Leadership and Followership. *Personality and Social Psychology Review* 10(4): 354-371.
Wrangham, Richard. 1999a. Is Military Incompetence Adaptive? E*volution and Human Behavior,* 20(1): 3-17.
Wrangham, Richard. 1999b. Evolution of Coalitionary Killing. *American Journal of Physical Anthropology* 29: 1-30.

Chapter 6: Oxytocin and the reduction of aggression

By Paul J. Zak, PhD
Professor and Director of the Center for Neuroeconomics Studies
Claremont Graduate University
Claremont, CA 91711-6165
paul.zak@cgu.edu

Aggression by humans has a variety of neurologic causes, including brain lesions especially in the orbitalfrontal cortex and amygdala; genetic variants, for example in monoamine oxydase A (MAOA) combined with an adverse developmental history; and variations in neurotransmitters and neuroactive hormones such as serotonin, testosterone and arginine vasopressin (Meht, Goetz & Carrè, 2012). More generally, episodes of aggression, especially repeated aggression by the same individual, are due to combinations of, and interactions between, genes, brains, history, and environments. Indeed, aggression on the field of play, or among soldiers toward enemy combatants, is promoted and acceptable, while abuse of spouses or random killing is inappropriate. This shows the situation-specific role of aggression that determines if it is warranted and acceptable or not.

While many social scientists view aggression (physical or with resources) as the norm, neuroscientific studies of human behavior--many from my lab--suggest the opposite: aggression is a useful but costly strategy and most people have a strong bias to cooperate in many situations. While aggression, and its cousin fear, are fairly easy to induce in laboratory settings, paradigms to study the neurobiology of cooperation began to emerge in the early 2000s. Many studies have used a neuroeconomic approach in which money could be sent to a stranger in the lab in a variety of settings to measure virtuous behaviors such as trust, trustworthiness, and generosity, as well as their absence, distrust, a lack of reciprocation, and greed. Understanding the positive side to human nature is valuable, these studies have shown, because it provides a richer neural and behavioral depiction of why neurologically healthy humans can alter their

behavior from cooperation to conflict. Conflicts could be as minor as yelling at a colleague at work, but may also encompass violent mass actions such as terrorist attacks. My own work on the neurobiology of moral behaviors has focused on the role of the neuroactive hormone oxytocin (OT). As I'll discuss below, OT appears to function as a chemical regulator that mediates prosocial behaviors by signaling that another person is safe or familiar, even if the other person is a stranger. While OT interacts with a host of other neurochemicals to affect behavior, its value in explaining human cooperation has recently been appreciated as evidenced by the number of recent books on the subject (Zak, 2012; Young & Alexander, 2012; Kuchinskas, 2009; Taylor, 2002).

OT, perhaps due to its ancient mammalian lineage, has several peculiar properties. It is one of the few hormones that is directly synthesized in the brain (like a neurotransmitter). It functions both as a hormone (has effects on the peripheral nervous system) and a neurotransmitter (is released into synapses in the brain). It is synthesized within a second or less of a stimulus, and has an approximately three-minute half-life, functioning much like an on-off switch signal safety. Lastly, and conveniently for experimentation, under physiologic stress animal studies have shown that the synthesis of OT by hypothalamic neurons coordinate central (brain) and peripheral (body) OT release. This means that an acute change in OT in the body is correlated with such an acute change in central OT. Yet, until a decade ago, OT was only studied in humans for its role as a hormone in reproduction (sex, birth, and breastfeeding).

There are several reasons OT was not studied in humans outside of reproduction. First, there is no medical disorder other than preterm labor known to be associated with too much or too little OT to prompt its study. Based on recent findings, though, there are now clinical trials for OT examining its role in the impaired social behaviors found in autism, social anxiety disorder, and schizophrenia. Second, OT is a "shy" molecule in that it has a short half-life and degrades rapidly at room temperature. When I began these studies in 2001, I had to develop tight handling protocols to capture the OT signal when it appeared and to minimize signal degradation. Third, although findings for the role of OT in promoting social behaviors in animals began to accumulate in the 1990s, most scientists had not found a behavioral task that would allow a test of the presumed prosocial effects of OT in humans.

Because humans appear to have more OT receptors in the forebrain than other mammals (Loup et al., 1991), and forebrain OT receptors modulate mid-brain dopamine circuits that reinforce and reward behavior (Zak, 2012; Donaldson & Young, 2008), one can make the case that cooperative behaviors are just as "natural" as aggressive ones. That is, the brain reinforces prosocial behaviors, revealing its value to the organism. Further, because recent studies have shown that OT is released even when strangers signal that they are safe and want to cooperate, a case could be made that cooperation with strangers is a typical human behavior, and that conflict among strangers may not be the norm.

Trust

The first nonreproductive stimulus in humans shown to induce OT release was a monetary transfer task known as the "trust game." In this task, strangers are seated in partitioned computer stations and all participants receive a $10 endowment for volunteering to be in an experiment. Identities are masked by using alphanumeric codes, and there is no deception of any type. Participants log in to computers and are randomly matched with another participant in the lab. The software randomly assigns participants to the role of decision-maker 1 (DM1) or decision-maker 2 (DM2). All DMs receive by computer the following instructions: DM1 will be prompted to transfer from $0-10 from his or her account to the DM2's account. Whatever is transferred is removed from DM1's account but is tripled in DM2's account. DM2 will receive a message through the software identifying the amount sent and the total in his/her account. DM2 is then prompted to send some amount of money, from zero to the total in her/his account to the DM1 who initially sent money.

This task had been designed by experimental economists (Hoffman, Dickhaut & McCabe, 1995; Smith, 1998) and had been run for both small and large stakes around the world. The consensus view in economics was the transfer from DM1 to DM2 was a measure of trust. Note it is not altruism or fairness that motivates a transfer since both DMs have the same amount of money initially. Yet, once DM1 transfers money to DM2 (about 90% of DM1s do this), DM2 now has entered into an implicit contract with DM1 that states "I trusted you because I believe you will reciprocate." Indeed, 95% of DM2s who receive money in this laboratory paradigm show they are trustworthy by reciprocating (Zak et al., 2004, 2005; Zak 2005). On average in these experiments, DM1s earn approximately $14 and DM2s earn $17, so their model of human beings as reciprocating creatures is, on average, correct. But why?

By taking blood after participants made decisions, my collaborators and I found that the more money someone received denoting trust, the larger the spike in OT. Further, OT in DM2s predicted how much money would reciprocate (Zak et al., 2004; 2005). The trust game captures in an objective way the notion of the Golden Rule: if you are nice to me, I'll be nice to you. Among the hundreds of people I have tested over the last decade in a number of variants of this task in a variety of cultures, roughly 95% of individuals reciprocate trust (Zak, 2012). The Golden Rule exists in every culture on the planet and reveals our essential social nature. It appears that OT is largely responsible for reciprocation by sending a safety signal motivating nice with nice (additional details in Zak, 2011).

One way to think about OT is that this molecule that evolved to facilitate live birth and motivate care for offspring in mammals is hyperactive in humans so that we often treat strangers like family. Since this is true for safe and stable environments for most people, the OT system allows us to quickly size up strangers and when appropriate derive value from relationships by cooperating with them. This also builds one’s reputation as a cooperator which is valuable for future interactions. Because synthetic OT is available and safe to give to humans, we tested whether if we manipulated the OT brain circuit pharmacologically we could induce greater trust in the monetary transfer task. Not only was trust increased for those infused with OT (via the nose), but we more than doubled the number of people who showed maximal trust by transferring all their money to a stranger in these experiments (Kosfeld et al., 2005).

These two sets of studies taken together showed that i) being trusted causes the brain to release OT and motivates reciprocation, and ii) exogenously increasing OT in people causes trust to increase. We showed the causal circle was complete. Nine other neuroactive hormones tested for their effect on trusting behaviors or OT release did not mediate these effects (Zak et al., 2005).

You will note that the trust game provides a win-win opportunity for participants, both can be made better off. In experiments with a win-lose task (more for you means less for me), OT increased generosity but only when the decision-maker had to take the perspective of the other person (Zak, Stanton & Ahmadi, 2007). OT also substantially increased donations to charity when the cause is made highly (Barraza & Zak, 2009) or minimally (Barraza, McCullough, Ahmadi & Zak, 2011) salient.

Pathology

Among the large number of studies my lab has run on OT, we consistently find that five percent of participants do not release OT when others do for a variety of stimuli. Investigating these individuals, we found they had some of the traits of psychopaths: sexual promiscuity, job instability, deception, and even self-deception (Zak, 2005). Because they do not release OT for positive social stimuli, I have coined the term Oxytocin Deficit Disorder (ODD) to describe them. Interestingly, their baseline OT is often very high. This indicates that their OT system is not processing social information in a safe/nonsafe way that others do. It also suggests a possible dysfunction with their OT receptors that regulate OT synthesis through a feedback loop. A recent study showed that those with diagnosed social anxiety disorder also appear to have ODD (Hoge, Pollack, Kaufman, Zak, & Simon, 2008). We have recently begun studying a large set of diagnosed psychopaths to explore the functioning of their OT systems in more detail.
Most psychopaths are identifiable as children or young adolescents (Kiehl, 2006), suggesting a strong genetic component. But we have also studied an acquired pathology due to repeated and severe sexual abuse suffered during childhood in a female clinical population. In a small sample of these patients that we intensively studied, we found roughly half of them do not release OT when shown trust. They also have impaired social behaviors, and particular difficulty modulating their behavior to the people or situation they are around. The majority of these patients were diagnosed with borderline personality disorder, were clinically depressed, and had psychosomatic medical symptoms (Zak, 2012). A sample of psychiatrically healthy women who had only suffered a few episodes of sexual abuse as children had intact OT systems and healthy social behaviors, for example, the ability to sustain fulfilling romantic relationships and family relationships. In the clinical sample, the severity and degree of abuse did not predict if they had acquired ODD. Resilience to abuse was predicted weakly by the presence of several genes, including genes that affect synaptic serotonin levels, though the sample was too small to have confidence in this finding.

Environment.

Multiple neurotransmitters activate multiple brain circuits to guide appropriate social behaviors as situations change from safe to unsafe. To simplify the discussion, I will focus on three, OT, testosterone, and epinephrine. Epinephrine (also called adrenaline) is the body's fast-acting stress signal. If something important, threatening, or difficult is occurring, epinephrine release will increase heart rate and respiration and prepare the organism to engage. Epinephrine is also an effective OT inhibitor (Jezová, Juránková, Mosnárová, Kriska, & Skultétyova, 1996).

Testosterone is also an OT inhibitor (Arsenijevic & Tribollet, 1998). If instead of "you are playing nice, so I'll play nice", the social environment shows that "you are playing bad", then testosterone increases to result in "I'll play bad back." This effect tends to be stronger in men who have five to ten times more testosterone than do women. We first found that men reciprocate bad with bad by turning the trust game on its head and asking what happens physiologically when people are distrusted, i.e. when they receive a small or no monetary transfer as DM2 in the trust game. We could not find a neural signal of distrust in women. But, in men the greater the signal of distrust, the higher the level of the "high octane" version of testosterone known as dihydrotestosterone (DHT). DHT levels spiked with distrust, and men with high DHT reciprocated little or no money to the DM1 who distrusted them. Women, on the other hand, were proportional reciprocators whether as DM2s they received small transfers or large ones (Zak, Borja, Matzner, Kurzban, 2005). Women did not have the "hot" physiologic response associated with distrust that men did. Note that in these experiments all DMs are anonymous so the gender of the DM1 is unknown.

To confirm this finding, we administered synthetic testosterone or placebo to 25 men and had them come to the lab twice in a blinded within-subjects design (once to receive testosterone, once to receive placebo). Using a zero-sum variant of the trust game that includes a costly punishment option, we found that men on testosterone, compared to themselves on placebo, were less generous in sharing resources with strangers, but more demanding of generosity from others (Zak et al., 2009). Indeed, these "alpha males" were more likely than their unenhanced selves to burn their own resources in order to punish others who had not cooperated. This could be the basis for establishing reputation or dominance in social relationships. Other studies have found that even the threat of punishment substantially increases cooperation (Boyd & Richerson, 1992). Men appear to bear the burden of punishment more than do women.

Environments that are unsafe, new, competitive, aggressive, or unpredictable can induce greater epinephrine and/or testosterone release and thereby inhibit prosocial behaviors, especially such behaviors towards strangers. Conversely, in environments that are familiar and safe, people have the luxury of releasing OT more often, possibly improving their family relationships, friendships, and opportunities to engage with strangers. I have collected evidence in my book *The Moral Molecule: The Source of Love and Prosperity*, that such environments can sustain a virtuous cycle of OT release, empathy, trustworthiness, and

happiness. Countries that are trustworthy have increased private investment in new businesses, creating jobs and reducing poverty (Zak, 2008; Zak & Knack, 2001). This permits a greater number of people to enjoy social connections and the OT release it potentiates, may stimulate greater virtue, mostly peaceful social relationships, benign international relations, and prosperity (Zak & Kugler, 2010). Quite a neat trick for an ancient molecule that was until recently largely ignored.

References

Arsenijevic Y, Tribollet E (1998) Region-specific effect of testosterone on oxytocin receptor binding in the brain of the aged rat. *Brain Res* 785: 167–170.

Boyd, R. Richerson, P.J. 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, 13(3), 171–195.

Donaldson, Z.R., Young, L.J., 2008. Oxytocin, vasopressin, and the neurogenetics of sociality. *Science* 322, 900–904.

Loup, F., Tribollet, E., Dubois-Dauphin, M., Dreifuss, J.J., 1991. Localization of highaffinity binding sites for oxytocin and vasopressin in the human brain. An autoradiographic study. *Brain Res*. 555, 220–232.

Hoge, E.A., Pollack, M.H., Kaufman, R.E., Zak, P.J., Simon, N.M., 2008. Oxytocin levels in social anxiety disorder. *CNS Neuroscience & Therapeutics* 14, 165–170.

Jezová, D., Juránková, E., Mosnárová, A., Kriska, M., & Skultétyová, I. (1996). Neuroendocrine response during stress with relation to gender differences. *Acta Neurobiologae Experimentalis, 56,* 779-785.

Kiehl, K.A. 2006. A cognitive neuroscience perspective on psychopathy: Evidence for paralimbic system dysfunction. *Psychiatry Research*, 142, 107-128.

Kuchinskas, S. 2009. *The Chemistry of Connection: How the Oxytocin Response Can Help You Find Trust, Intimacy, and Love*. New Harbinger Publications

Smith, V.L., 1998. The two faces of Adam Smith. *Southern Economic Journal* 65, 1–19.

Taylor, S. 2002. *The Tending Instinct: How Nurturing is Essential to Who We Are and How We Live*. Henry Holt & Co.

Young, L., Alexander, B. 2012. *The Chemistry Between Us: Love, Sex, and the Science of Attraction*. Current Hardcover.

Zak, P.J. 2012. *The Moral Molecule: The Source of Love and Prosperity*. New York: Dutton.

Zak, P.J. The Physiology of Moral Sentiments. (2011) *Journal of Economic Behavioral and Organization*, 77(1): 53-65

Zak, P.J., Kugler, J. 2010. Neuroeconomics and international studies: a new understanding of trust. *International Studies Perspectives*, 12, 136-152.

Zak, P.J., Kurzban, R., Ahmadi, Swerdloff, R.S., Park, J., et al., 2009. Testosterone administration decreases generosity in the ultimatum game. *Public Library of Science ONE* 4, e8330, doi:10.1371/journal.pone.0008330.

Zak, P.J., 2008. The neurobiology of trust. *Scientific American*, June, 88-95.

Zak, P.J., Stanton, A.A., Ahmadi, S., 2007. Oxytocin increases generosity in humans. *Public Library of Science ONE* 2, e1128.

Zak, P.J., Borja, K., Matzner, W.T., Kurzban, R., 2005a. The neuroeconomics of distrust: sex differences in behavior and physiology. *American Economic Review Papers and Proceedings* 95, 360–364.
Zak, P.J., 2005. Trust: a temporary human attachment facilitated by oxytocin. *Behavioral and Brain Sciences* 28, 368–369.
Zak, P.J., Kurzban, R., Matzner, W.T., 2005b. Oxytocin is associated with human trustworthiness. Hormones and Behavior 48, 522–527. Zak, P.J., Kurzban, R., Matzner, W.T., 2004. The neurobiology of trust. *Annals of the New York Academy of Sciences* 1032, 224–227.
Zak, P. J., Knack, S. 2001. Trust and Growth. *The Economic Journal* 111: 295-321.

Part III: Systems Understanding

Toward a Systems Continuum: On the Use of Neuroscience and Neurotechnology to Assess and Affect Aggression, Cognition and Behavior

James Giordano PhD, MPhil
Division of Integrative Physiology, Dept of Biochemistry
and
Neuroethics Studies Program
Georgetown University Medical Center, Washington, DC, USA
and
Human Science Center
Ludwig-Maximilians Universität, Munich, GER

Roland Benedikter PhD, DrPhil
The Europe Center
Stanford University
Stanford, CA, USA

Introduction: Advances in Neuroscience and Neurotechnology

As can be seen from the diversity and depth of contributions to this whitepaper, neuroscience has assumed a progressively prominent role in shaping views of the human being, human condition, and human relationships. In this light, there is a strong – and we believe defensible – sentiment within the scientific, military, political and public communities, that the brain represents the "next frontier" of scientific exploration, discovery and intervention. Through the use of iteratively more advanced techniques and technology, neuroscience - or perhaps more accurately, neuroscience and technology (" neuroS/T") has enabled an enhanced understanding of nervous systems on a variety of levels. At present, we have a generally solid working knowledge of the substrates and mechanisms of neurological structure and function,

respectively, what neural cells and networks are made of, and the activities of these cells and structures. However, we have just begun to extrapolate this to a fuller conceptualization of brain function as a complex, dynamic, system,, and how neural systems are affected by – and affect - natural systems, at-large (von Bertalanffy 1968; Schoner and Kelso 1988; von Weizsacker, Lovins, and Lovins 1998; Juarrero 2002)

Neuro-ecology: Interacting Systems of Neurobiology and Culture.

Current neuroscientific perspectives consider biological organisms to be complex (internal environmental) systems nested within complex (external environmental) systems (Schoner and Kelso 1988; Juarrero 2002). Interactions within and among systems are based and depend upon numerous variables within these internal and external environments (Ridley 2003). Given the definition of *ecology* as a study or system of reasoning about the interrelation of organisms in their environment or place of inhabitance (Merriam-Webster's Collegiate Dictionary 2004), we argue for consideration of neuro-ecology (i.e.- the neuroscience of human ecology) as "studies of interactions between neural systems embodied in individuals, that are embedded in groups and environment(s) framed by time, place, culture and circumstance" (Giordano 2011a; Giordano 2011b; Giordano, Benedikter, and Kohls 2012). This mandates appreciation of culture as an important force in determining the interactively neural-cognitive/emotional-environmental (i.e.-bio-psychosocial) dimensions of human functioning. At the most basic level, *culture* refers to a medium for the development of living material, and it becomes important (if not necessary) to evaluate how "culture" engages and sustains the set of shared material traits, characteristic features, knowledge, attitudes, values and behaviors of people in a common place and/or time. This definition rightly reveals that culture establishes and reflects particular biological characteristics (that develop, and are preserved in response to environments), that can be expressed through cognitions and behaviors. In this way, culture is a medium for bio-psychosocial development, and a forum and vector for its expression and manifestations (Ridley 2003; Giordano, Engebretson, and Benedikter 2008). *Defining the neural bases of such biological-environmental interactions may yield important information about factors that dispose and foster various actions - including cooperation, conflict, aggression and violence* (for overviews, see: Cacioppo, Visser, and Pickett; Verplaetse, DeSchrijver, Vanneste, and Braeckman 2009).

On a variety of levels, , neural systems allow individual agents/actors to intuit, relate, and react/respond to the multiply tiered environments in which they are nested; a simplified depiction of this relational activity is schematically presented in Figure 1.

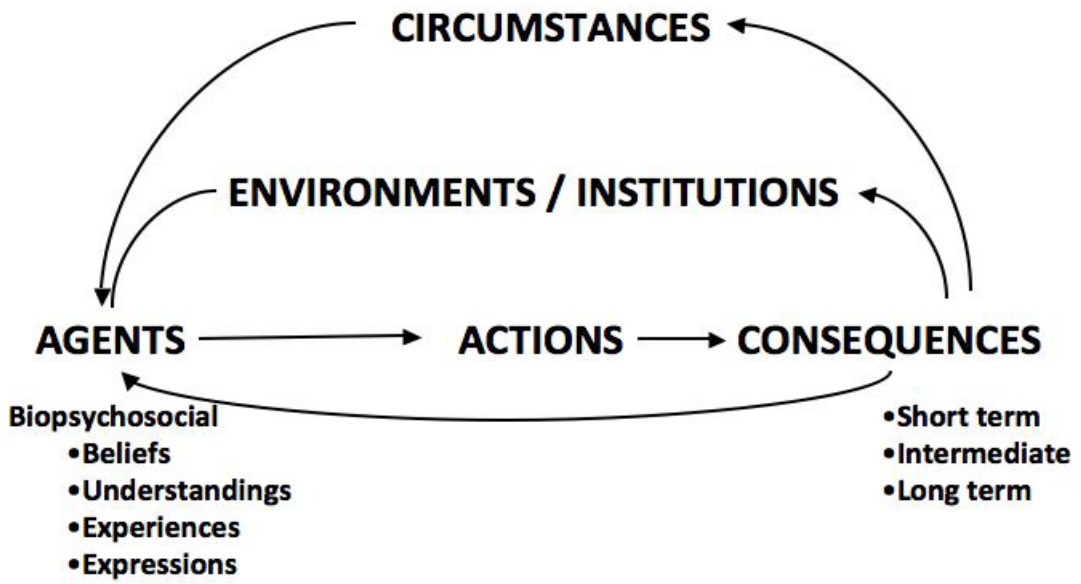

Figure 1. Interacting bio-psychosocial (i.e.- ecological) domains of environments, agents, and actions. Arrows indicate feed-forward and feed-back potential and patterns that incur and sustain complex dynamical (and cybernetic) properties of the system. Note that features of the biological domains (e.g.- population and community genomes; individual and group genetics; individual and group phenotypic expressions of structure and function), psychological domain (e.g.- beliefs, understandings, experiences, expression) and environmental (consequential) domain (e.g.- short, intermediate, and long-term outcomes) interact (bi-directionally) within (i.e.- vertically) and between (i.e.- horizontally; viz.- dispositionally and consequentially) these domain regions, to create clades or bio-psychosocial trend-patterns within particular ecological settings (of place, situations and time).

Neural systems function in decision-making behavior(s) by enabling orientation of the present to recollection of, and relation to the past, in order to anticipate/predict future outcomes and consequences (based upon extant predispositions and prior experiences), so as to influence and determine (current) attitudes and actions. This can be simplified (to a considerable extent) and summarized as a modification of Boyd's *Observe, Orient, Decide, Act (OODA)-schema* (Boyd 1995), to represent a neuro-cognitively dynamic loop inclusive of, and responsive to dispositions and consequences. There is a tendency toward Bayesian functions, in that prior experiences and activities of component networks within the system (i.e. - consequences) create "weighted" or biased patterns of neural network activity (i.e.- dispositions) that are hierarchically expanded into patterns of cognitions, emotions and behaviors. This process is schematically illustrated in Figure 2.

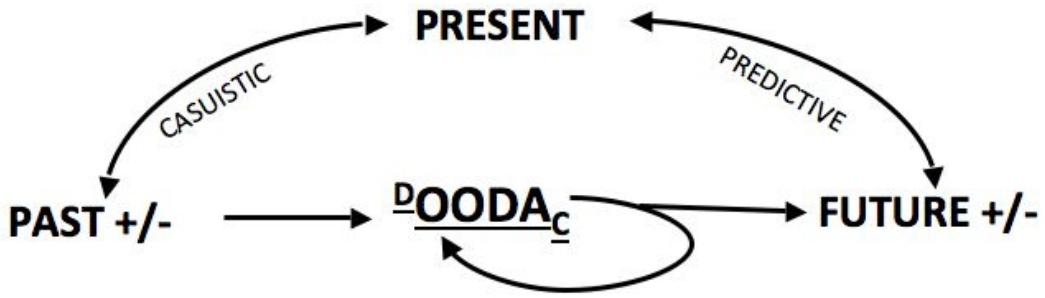

Figure 2. Putative neuro-cognitive systems' dynamics relating past, present and future experiences relative to decision-making. Note that the (OODA) loop is modified by previous dispositions (D) and resultant consequences (C) that are manifest both internally and externally within and across domains of the system.

These constructs have provided a basis for understanding "neuro-ecology", which we propose as a substantive framework for connecting basic neurobiological activity, to socio-cognitive function, to external environmental and cultural milieu. We offer that this framework represents a superset within which a similar framework for "neurodeterrence" could be considered

Neuroscience has become a convergent discipline, engaging techniques and technologies from the natural and physical sciences, as well as the humanities to address questions of psychology, and more frequently, sociology, anthropology and economics (Wilson 1998; Giordano 2012a; Giordano, Benedikter and Flores, 2012; see also prior discussions in this white paper). Neuroscientific discoveries are fostering re-examination of, and may challenge socially-defined ontologies, social values, conventions, norms and mores, and the ethico-legal notions of individual and social "good". (Giordano 2011a, Giordano 2011b; see also: Rees and Rose 2004; Illes 2005; Rose 2005; Glannon 2007; Giordano and Gordijn, 2010). Given the rapid development of evermore sophisticated neurotechnologies (e.g. progressively more capable iterations of neuroimaging, brain implants and brain machine interfaces; neurogenetic and tissue transplants; trans- and intracranial stimulation; etc; see: Giordano 2012c for overview), it is vital to ask how these will be used to assess, target and control the cognitions, emotions, and behavior of individuals, groups and potentially, even societies.

We can only speculate about the possible ways that consciousness (that is, the function colloquially referred to as "mind"), could occur in brain, and how neurophenomenological cognitions and emotions influence biology and behaviors - what philosopher David Chalmers' classifies as one of the principal "hard problems" of neuroscience (Chalmers 1995). Despite

these problems, neuroscience is being increasingly employed to assess and affect thought, feeling, behaviors, and more broadly, constructs of normality and abnormality (for review, see Jeannotte et al. 2010; Swaab 2010; Giordano and Du Rousseau 2011). As technology advances and neuroS&T becomes more readily available, it will be used in the national security contexts broadly, and, we believe, more specifically in neurodeterrence.

In light of this, the following sections of this chapter describe and examine key issues that must be considered and assessed in order to apply neuroecological models to potential tools that can be used in national security. Many of these same considerations are relevant to and directly applicable to the framework - and employment - of neurodeterrence, as this concept and field develops in similar contexts and trajectories

NeuroS/T to Assess and Affect Human Ecology

To date, most efforts toward global relations, and security and defense have focused upon social and cultural factors influencing a host of human behaviors, including patterned violence and terrorism. Given that these behaviors are expressed by human actors, and humans are bio-psychosocial organisms embedded within, and responsive to, geo-cultural environments, then we offer that it is important to address and discern those (neuro)biological, psychological and social factors that instigate violence. This emphasizes the viability and potential value of neuroS/T in programs of national security, intelligence and defense (Giordano, Forsythe, and Olds 2010; Forsythe and Giordano 2011), and we opine that appropriate use of neuroS/T in security operations (i.e.- "neurosecurity") is (and will be increasingly) influential to deterrence and defense. The challenges posed for using neuroS/T in these ways are: 1) to develop a more complete understanding of mechanisms that precipitate aggression and patterned violence; 2) to provide practical and ethical options to affect, alter and/or impede these mechanisms, and 3) to base any such findings, options and actions upon realistic appraisal of the capability, limitations and hence, practical, ethico-legal and socio-political direction, and in some cases, constraint of this science, technology and information.

Neurosecurity - A Key Component of Deterrence and Defense

Herein, we provide a 2-fold definition of *neurosecurity* as studies and applications of: (I) the concepts, practices, guidelines and policies dedicated to (a) identifying socio-political and military threats to neuro-psychiatric information and function, and (b) preserving the integrity of both neuro-psychiatric information and neuro-psychiatric function of persons, groups and populations; and, (II) neuroscientific techniques and neurotechnologies to affect, manipulate and/or control neurological structures and/or functions of individuals, groups and/or populations in the service of national defense, and/or military objectives. As history illustrates,

new developments in science have- and will continue to have - particular appeal for use in security and defense agendas, and this is certainly the case for neuroS&T (Bitzinger 2004).

A 2008 report conducted by the ad-hoc Committee on Military and Intelligence Methodology for Emergent Neurophysiological and Cognitive/Neural Science Research in the Next Two Decades National Research Council of the National Academy of Sciences entitled "*Emerging Cognitive Neuroscience and Related Technologies*" addressed the state of neuroscience as relevant to the (1) potential utility for defense and intelligence applications, (2) pace of progress, (3) present limitations, and (4) threat value of such science (National Research Council 2008). Stating that "…military and intelligence planners are uncertain about the likely scale, scope, and timing of advances in neurophysiological research and technologies that might affect future U.S warfighting capabilities" (National Research Council 2008, 14) the Committee essentially defined the state of the field in its assertion that "…for good or for ill, an ability to better understand the capabilities of the body and brain will require new research that could be exploited for gathering intelligence, military operations, information management, public safety and forensics." (National Research Council 2008, 14). As the Committee report noted, there is a fair amount of "… pseudoscientific information and journalistic oversimplification related to cognitive neuroscience," (National Research Council 2008, 3) and so any consideration of the possible use of neuroS&T for national security, intelligence and defense (NSID) would need to parse facts from fiction about what these approaches actually can and cannot do. The goal is not to be dismissive, but rather to be critically perceptive, and keen to the potential for innovation and viable ways that neuroS&T could be developed, used and/or misused, to what ends, and by whom.

Simply put, the brain and nervous system can – and *will* – be engaged to effect outcomes relevant to NSID operations, and some of these efforts will most certainly be undertaken by countries other than the United States and its allies. (Giordano, Forsythe, and Olds 2010) Thus, it is crucial to remain keenly aware of international research programs that could be used in ways that pose obvious threat(s) to security and defense. Surveillance of international research, development, testing, and evaluation (RDTE) is necessary, but insufficient to guard against such potentially negative and harmful uses of neuroS&T. Instead, we advcoate a stance of national/public security that is based upon preparation, resilience, and in some cases intervention, to prevent the advancement of certain RDTE trajectories (as well as other potential threats of aggression and violence).

This will require the coordinated discourse between scientists, engineers, ethicists, sociologists, futurists (forecasters?), and the public (with diligent stewardship of information so as to balance relative transparency and vulnerability of sensitive details relevant to the integrity of national security). The discourse should conjoin academic, corporate, and governmental sectors (the so-called 'triple helix' of the scientific estate; Etzkowitz 2008) at a variety of levels and stages in this enterprise (Wurzman 2010; Wurzman and Giordano, 2011). This is not new; we need only to look at the Manhattan Project and 'Space Race' for examples of this estate in practice (Etzkowitz 2008). But that framework, while viable, may require modification(s) to

facilitate the type and extent of convergent approach that allows for stronger collaboration between the (physical, natural and social) sciences and the humanities (Giordano 2012a).

We argue that this involvement of both the humanities and the public (at least to some reasonable extent) is important because any real effect – both domestically and internationally – can only be leveraged through guidelines, laws and policies that are sensitive to ethical and social effects, issues, and problems. But international policies don't guarantee cooperation (Gregg 2010). So, any meaningful efforts in neurosecurity must sustain an active research program that delves into the potential capabilities and limitations of neuroS&T, and enables ongoing evaluation of possible future S&T applications. The axiomatic goal of national security is the protection of the population. Toward this end, knowledge of real and potential threats is crucial to both preventing events that place the population at risk, and to mitigate events before they escalate into scenarios of large-scale harm. Intelligence is a vital part of any national security agenda, and accurate information is the key to successful intelligence (Davies 2010). There is increasing interest – and concern about – developing and using neuroS/T to enable more effective intelligence in domestic and international settings; both of which may present complex cross-cultural issues and problems.

Techniques and technologies that have been identified as having possible utility for obtaining information that could be important to intelligence efforts include:
1) a variety of neuropharmacologic agents, including substances that induce feelings of affiliation , mood altering drugs (such as the anti-anxiety drugs and dopamine transport inhibitors) and drugs that produce a state of elation or euphoria (such as some of the opiates, and amphetamines, e.g.- methylenedioxymethamphetamine, MDMA; for overview, see elsewhere in this whitepaper, and Wurzman and Giordano 2011); and
2) neurotechnologic devices and approaches, such as certain types of neuroimaging, and forms of magnetic and\or electrical nerve and brain stimulation (e.g.- transcranial magnetic stimulation, TMS; Wurzman and Giordano 2011)).

Dismissing the possible employment of neuroS/T, based upon either fear of misuse and/or ethical qualms, does not reflect the historicity of using state-of-the-art S/T in NSID agendas, and therefore we argue that such a view may be unrealistic. However, we also believe that it is necessary to establish three important premises about the use of neuroS/T in NSID: First, it is likely that in the near future, neuroS/T will become (more) widely used in intelligence gathering and implementation of security and defense. Second, neuroS/T, like any scientific approach and tool, has potential for misuse, and so identifying the nervous system and brain as target sites through which to incur frank harm(s) in the acquisition and leveraging of information and the modifications of emotions and behavior is a reality that must be faced. Third, it's probable that other individuals and/or groups are also focusing upon these goals and tasks, and such intent may not be friendly to the US and its allies.

From these premises, we offer a three-pronged stance: First, a realistic acknowledgement of the actual capabilities and limitations of the neuroS/T used, and the ethico-legal issues generated by apt or inapt use, or blatant abuse – is required. Second, is the

need to avoid the so-called *fallacy of two wrongs* (Groarke 1982), and not simply 'do something' (or do something cavalierly or without appropriate reflection and regard) just because "…someone or everybody else might". Third, is the additional need to be prepared for the contingencies and realties of such uses of neuroS/T, but equivalently, to do so in ways that are scientifically and technologically apt, and ethico-legally sound (Giordano 2012c; Giordano and Benedikter 2012a).

To bolster this stance, we (Bower and Giordano 2012; Benedikter, Giordano, FitzGerald 2010) have posed the following questions as defining and shaping the conduct of neuroS/T research and use:

- Is there some "sanctity of mind" (Fields 2006) that negates the use of such approaches, regardless of how suspect an individual may be?
- Or, are there particular circumstances under which certain advanced neuroscientific methods may be employed to obtain intelligence and incur deterrence in light of real and significant danger to the populace?
- Does the use of neuroS/T incur *greater* or *lesser* risk and harms than other intelligence, security, deterrence and defense methods?
- Are there limits to the ways that neuroS/T should be used in such situations, and if so, how should such criteria be developed and enforced?

Of course, there are claims that neuroscientific methods should not be employed in interrogation – or national security agenda at all – because of the potential for misuse, and/or the view that using neuroS/T in these ways would incur violations of inherent human rights that the US and its allies have vowed to protect (Bell 2010; Benanti 2010).

We recognize and respect the validity of such claims, and in light of this we envision three possible options:

1. Abstaining from implementing neuroS/T in any/all national security agendas and situations.
2. Utilizing neuroS/T in only specific situations/conditions that would dictate – and ethico-legally justify – the need for this level of intervention.
3. Making (appropriate) neuroS/T approaches available and employable in all national security endeavors, including interrogations, in accordance with defined ethico-legal parameters.

When considering these options, it is important to bear in mind that the appointed goal of intelligence for US (and NATO) national security and defense is not to cause harm without purpose, but rather, to uphold and protect the rights of the greater population (namely the right to life; Gross 2001). But, as history has shown, law enforcement and military authority can be misappropriated and abused, and these possibilities must be taken into account and mitigated.

In the main, we call for a focus upon:

1. Whether to base ethical decisions upon the spirit of the law, which might allow such uses of neuroS/T (Montesquieu 1978); or, if such approaches would be considered so morally problematic that it would be preferable to ban the development and implementation of these techniques and technologies altogether;
2. Whether guidance and governance should entail a neuroethics of military operations – or a military ethics applied to the use of neuroscience and neurotechnology (Bower and Giordano 2012);
3. Whether some (extant or new) combination of both approaches might need to be addressed and articulated, and what such a set of ethico-legal parameters would obtain and entail.

We advocate that neuroS/T be continued to be studied for its potential viability – specifically to *decrease* harms necessary to preserve national security and defense. However, we urge sensitivity to what we call "*footfall effects*": namely, that it is not a question of impeding the momentum or even the pace of forward progress (because that may be difficult, if not impossible, to do); rather it is a question of where each forward step falls, so as to tread wisely with appropriate lightness or force, and remain upright and balanced, both in the course of usual events, and if pushed or stricken.

Clearly, there are a number of issues, problems and concerns that come to the fore, but at this point, we are focused upon two: First is whether such neuroS/T is mature enough to be used in these ways. An expanding body of literature supports the use of neuroscientific techniques and tools to provide new insights into how cognitive and emotional systems could be manipulated to affect the perception of the past, present, and future. While this may allow utilization of neuroS/T it is should be noted that each and all of these approaches possess particular capabilities and limitations.

For example, neurogenetics and neuroproteomic assessments can provide detailed information about neural predispositions, and the presence of neural biomarkers that have been putatively associated with, and may be inferentially predictive of particular cognitive, emotional and behavioral characteristics. Yet, it is well recognized that it is difficult – if not often erroneous - to attempt linear and/or direct correlation of populational genomic, and individual genetic and proteomic markers to psychosocial traits and states, given the complexity of single- and multiple-gene effects, and the ongoing dynamics of genetic-phenotypic, and environmental interactions in shaping psychosocial outcomes (Wurzman and Giordano 2012). Various types of neuroimaging (such as computational tomography, CT; functional magnetic resonance imaging, fMRI; and diffusion tensor imaging, DTI) provide generally good spatial resolution of regional activity in the brain; however, the temporal fidelity of these techniques leaves much to be desired. Neurophysiological techniques, such as quantitative electroencephalography (qEEG) and magneto-encephalography (MEG) have good temporal resolution, but tend to lack finely-grained spatial integrity (VanMeter 2010).

Many of these shortcomings can be de-limited through the convergent utilization of multiple forms of neuroscience and technology (e.g.- genomics and genetics; proteomics; neuroimaging; individual and group socio-behavioral analyses, etc), so as to provide an integrative montage or mosaic of information about neuro-cognitive predispositions and individual and group characteristics that may influence patterns of cognitions, emotions and behaviors (Giordano 2012a; Vaseashta 2012). The proposed use of such convergent neuroS/T approaches is to: 1) assess individuals from selected geographic and cultural regions; 2) create iterative data bases to develop comparative and normative inferences specific to characteristics of groups and populations within these geo-cultural domains; 3) employ these data to model neuro-biopsychosocial dynamics that might contribute to violence; 4) use these data, models, and norms to better define and predict individual and group behaviors, and 5) engage this understanding to mitigate factors that foster and/or initiate violence.

These neuroS/T approaches are not intended to be applied to all members of a given population; rather, it is critical to accumulate an amount and levels of data that are necessary and sufficient to extrapolate group comparisons and predictions. This necessitates employment of computational technologies (e.g. - large scale databanks, cloud computing) to afford the resources and services required to store, integrate and retrieve such information with accuracy and expedience. In the practical sense, such data could be utilized to provide indications for individual and/or group tendencies toward particular cognitive, emotional and behavioral trajectories, so as to indicate (and/or warrant) further, more finely-grained assessment of certain individuals or groups, and initiation of some form of mitigating interventions (Giordano 2012a).

This will require both ongoing assessment of the viability of extant neuroS/T, and continuing identification and analyses of gaps in information, capability and administrative structures to provide oversight of these current and emerging tools and techniques (Shaneyfelt and Peercy 2012). There is a real risk that neuroscientific outcomes and information may be misperceived, and misused to wage arguments that are inappropriate or fallacious. Misperception and/or misuse can result from miscommunication of what neuroscientific data actually mean. In this sense, we have advocated discernment of “hard” from “soft” neuroscience: the former being that which is actually produced and disseminated within the scholarly community, while the latter tends to be that which is excerpted, or in some cases, bastardized in the extra-academic sphere (Giordano 2011b). This speaks to the shared responsibilities of science and various user communities (including public media) to avoid ubiquitous flaunting, claims and/or demands of neuroS/T in ways that are nonsensical. As Matthew Crawford has claimed, the limits to “neurotalk” need to be recognized and appreciated (Crawford 2010).

This latter point prompts our second focus upon questions and concerns about whether ethico-legal systems are in place and realistic and mature enough to guide, direct and govern such possible use and/or non-use. In short, we claim that they are not; at least not to the extent that we believe necessary and sufficient to address and account for the contingencies spawned by rapid advancement in neuroS/T and the pull exerted upon its use and employment by

a variety of economic, social and political forces (Giordano, Forsythe, and Olds 2010; Forsythe and Giordano 2011; Giordano and Benedikter 2012a,b). This is where the proverbial "rubber hits the road" as regards the ways that pragmatic evaluations of the capabilities and limitations of neuroS/T are translated to practical parameters for the ways that these approaches can, should, and/or should not be utilized.

In light of this, we are attempting to develop algorithmic protocols for studying and using neuroS\T that:

1. Reflect and substantiate technical rectitude;
2. Reflect appropriate moral analyses of use and outcomes;
3. Afford ethico-legal bases to guide/direct both the use of neuroS\T and its outcomes within extant judicial frameworks and guidelines of international relations, security and deterrence; and
4. Engage technical and ethical concepts to revise/develop pertinent laws to ethico-legally govern any use of neuroS/T in such circumstances.

From these studies, we are developing a proposed set of criteria for using neuroS/T in national security settings (Bower and Giordano 2012). These tentative criteria include:

1. That there is less harm done by using the neuroS/T in question.
2. If an individual or individuals pose(s) a realistic and immediate threat of severe harm to others, the most effective science and technology – and least harmful among these – should be utilized toward mitigating these threats.
3. The use of such neuroS/T must be admissible in a court of law under Daubert (i.e. – reliability) rather than merely Frye (i.e. - relevance) standards (Orofino 1996). As well, we are examining other ethico-legal frameworks and standards to enable a more internationally relevant approach to using neuroS/T in such ways (see, for example: Eagleman 2011).
4. If neuroS/T is employed for intelligence purposes, only information pertinent to an ongoing investigation or a specific issue of security and/or deterrence should be obtained and used, and this should be stored in official police and/or government records.
5. There must be other corroborating evidence to substantiate prosecution and interventive action(s) -outside of evidence gathered by neuroS/T - as is necessary based upon maturity and reliability of techniques (see 3).
6. There must be a valid legal order issued to incur use of neuroS/T in these circumstances (see 2 and 3).
7. Applying these technologies in a preventive or predictive manner is still practically problematic and should not be implemented until further S/T research and development has been undertaken, and adequate ethico-legal frameworks are addressed and generated.

We are also working to develop policy recommendations that are aimed at supporting fiscal investment in building sustainable infrastructures that:

1. Engage research to evaluate if and how neuroS/T could be used in NSID

2. Develop a stance of preparedness with respect to the potential military and law enforcement uses of/for neuroS/T.
3. Establish multi-disciplinary bodies to formulate ethico-legal guidelines and protocols to monitor/oversee/regulate the use of neuroS/T both in the US, and internationally.

Practical Questions; Ethico-legal Concerns: Issues of Power

Of course, this paradigm generates both questions of the ecological validity and reliability of any such assessments, as well as ethico-legal concerns about the value and probity of predictive neuro-cognitive assessments to compel various forms of pre-emptive intervention. Without doubt, there is the need to develop stringent technical and ethico-legal guidelines and standards for such use of neuroS/T – a project to which our group remains durably committed. We posit that the challenge reflects, and must address important standing questions in the field. Namely, what are the nature and type of neurobiological characteristics that affect cognition, emotion and behavior? Can these characteristics be accurately assessed, and what types and combinations of techniques, technologies and metrics are required in this task? Can these, techniques, methods and tools – if not overall paradigm – be used to (a) describe and perhaps predict bio-psychosocial factors of group violence and terrorism, and (b) provide putative targets for multi-disciplinary intervention to deter, mitigate and/or contain such violence?

In the main, we warn against succumbing to "Icarus' folly" of scientific and technological hubris. Simply put, it is unwise – and inapt – to over- (or under-) estimate the capability of neuroS/T, and it is equally foolish to misjudge the power conferred by this science, or the tendency for certain groups to misdirect and misuse these technologies and the power they yield (Giordano 2012c). In light of this, we call for a concomitant dedication to both ongoing neuroS/T research, and full content ethico-legal address, analyses and articulation of the ways that these approaches may be used, misused and/or abused in contexts of national security, intelligence and defense (by the United States and its allies, as well as other nations on the world stage). Prescriptions, proscriptions, and guidelines must be devised and implemented to ensure the technically apt and ethically sound use – and governance - of such methods and information.

To be sure, there is robust political power to be gained and leveraged through the use of neuroS/T. Although reliant in part upon economics, such power transcends simple economic considerations. In the changing political constellation of today's broad "power shift from the West towards the East which is the consequence of the latest economic and financial crises" (as asserted by France's Premier Francois Fillon on 6 November 2011; Evans-Pritchard 2011), neuroS/T is becoming a crucial international factor. Given the growing prominence of non-Western nations in neuro- and biotechnology research and production, the adage that "the one who controls the chips controls the game" is metaphorically accurate in that these nations' efficient production of bio- and neuro-technologies are fostering a presence on the world stage, thus creating new social and political dependencies.

These developments suggest that neuroS/T is becoming a strong factor in the re-balancing of the power equation of global politics, influence, and defense capabilities. The issue of how to apply neuroS/T prompts the human question in the more strict philosophical and ethical sense, as reflective of, and inherited by Western modernity (and which will heretofore be situated amongst increasingly prevalent and influential pluralist ideals and ethics on the 21st century world stage). The question of what the human being is, and what it may become, is – and will be - unavoidably connected with the development, control and policies of neuroS/T. To re-iterate, neuroS/T - like any form of science and technology - can be used to effect good and harm. And while the tendency to use science and technology inaptly or toward malevolent ends is certainly not new, the extent and profundity of what neuroscientific information implies (i.e. about the nature of the mind, self-control, identity, and morality) and what neuroS/T can exert over these aspects of the human being, condition, predicament and relationships mandates thorough review and discernment.

Therefore, a particularly high level of scrutiny is needed when looking to, and relying upon neuroS/T, both for determination of ethico-legal judgments, and to describe, predict or control human behavior. It will be crucial to develop measurements for such scrutiny, and how to translate these metrics to binding legal standards nationally, regionally and internationally. Extant criteria, such as the previously mentioned Frye and Daubert standards used in the United States, while viable to some degree, are changeable, can reflect - and are often contributory to - the scientific, social and economic "climate" in which various techniques and technologies are regarded, embraced and utilized, and thus in most cases are only temporary political agreements, and (in the ideal case) socially accepted viewpoints. Thus, any and all analyses and guidelines for the use of neuroS/T must be based upon pragmatic assessment of technological and human dimensions of science and technology, the capabilities and limits of scientific and technological endeavour, and the effects and manifestations that studying and using such science and technology might incur in the public sphere.

Addressing Challenges and Opportunities: A Path Forward

We argue that it is essential to appropriately address these "deep" questions, both separately, and in their inter-relatedness. A first step is to more fully recognize the rapid development and use of neuroS/T, and the variety of new fields of application and transformation generated in the mid- to long-term by neuroscientific techniques and tools, and the information and capability they yield. To date, the US and its allies have not forged concrete political strategies and policies to optimize the beneficial effects of neuroS/T on the one hand, and confine potentially negative effects on the other. Instead, there appears to be a somewhat indecisive posture - a "waiting game" - toward the unavoidable increase in contextual (indirect) and classical (direct) political power connected to the use of neuroS/T in the coming decades.

We believe that what is needed is the formulation and articulation of a democratically defined, multi-disciplinary neuroethics. Such a comprehensive and cosmopolitan neuroethics

does not yet exist on an international global level, and thus the US and its allies could be at the forefront of proposing, developing, elaborating, implementing and promoting such ethical progress. We maintain that this type of neuroethics, brought forward as a new, inclusive strategy for the development of transnational neuroscientific innovation - on a global scale - will be instrumental to developing political, economic and military relationships (Giordano 2010; Giordano 2011a; Giordano and Benedikter; 2012a; Giordano and Benedikter 2012b; Giordano, Benedikter and Flores, 2012; Shook and Giordano 2013).

NeuroS/T is and remains a field in evolution. The questions generated by the field and its applications are complicated – and more numerous than the certainties achieved thus far. The common ground of these questions is not whether "deep reaching" scientific and technological shifts will occur, but rather when, and to what extent. And the most overarching question is how, and in which ways these shifts will be expressed by, and/or affect global political forces. Could some uniform regulations for research and use be viable in any and all situations? And if so, by which mechanisms might these codes be developed and articulated? Or, will progress in neuroS/T incur more of isolationist leanings? And, in the event, how would Western nations then maneuver neuroscientific efforts to retain a viable presence on the global technological, economic, security and defense map(s)? Might this trend toward pervasive use of neuroS/T in these silos of power be regarded as a form of "neuro-politics"?

It is exactly this scientific-to-social span of neurosS/T effects that necessitates a stronger focus and investment in both the science and a meaningful neuroethics. As a new, "proto-political" discipline, neuroethics entails and obtains two main traditions (Roskies 2002; Racine 2010; Giordano 2010; Giordano 2011a; Levy 2011). The first is focused upon the nature and patterns of human cognition, needs and resource utilization in moral, ethical and social decision-making. This approach is important to appreciate the ways that bio-psychosocial (including cultural) differences are manifest. Yet, insight and understanding of such putative substrates and mechanisms are not sufficient to foster inclusive political perspectives. Therefore, it will be necessary to engage this knowledge both in the analyses of problems borne of neuroscientific and neurotechnological progress, and the development of recommendations and guidelines that direct the scope and tenor of current neuroscientific research and applications, so as to ensure preparedness for the consequences of neuroS/T advancements in the future.

It may be that existing ethical and legal concepts and systems need to be adapted, or even developed anew to sufficiently account for the changes and challenges that neuroS/T are evoking in an evermore pluralized world (Giordano and Benedikter 2012a; Giordano and Benedikter 2012b; Giordano and Shook 2013). As philosopher Fritz Jahr noted some 80 years ago, new science and technology unavoidably add to the palette of philosophy, ethics, economy, culture and politics alike (Jahr 1927). They require new forms of ethical reflection, and revised concepts and enactments of policy. Given the growing reciprocal relationship of knowledge, technology, politics, economics and culture in the years ahead, in order to maintain leadership and influence on the global stage it will be necessary to develop guidelines, policies and laws that appropriately reflect advancements in neuroS/T, and aptly direct, if not govern

the use and manifestations of such developments in an era of an intensifying global systemic shift.

Conclusions

Our intent is not to advocate the use of any particular neuroS/T in national security, intelligence and defense, but rather to illustrate that any and all consideration and analysis of use must begin with and proceed from fact(s). The fact *is* that neuroS/T, like any science and technology, can and will be used in service of national security and defense agendas, not only by the United States, but by nation states and individual actors with aims that are not aligned with the values of the USA and its allies. This finding arises from our ongoing work in surveillance of the field and the ways that neuroS/T is – and potentially can be – used and misused. It is from this reality that we invoke a stance of preparedness and an ethic of responsible intent and action (Giordano, Forsythe, and Olds 2010). Given the trend toward revivifying US "Big Science" incentives on a global scale (inclusive of, if not explicitly focusing upon the brain sciences), and the increasing advancement of neuroS/T worldwide, we believe that such investment of time, effort and funding is important, necessary and urgent.

Acknowledgements

This work was supported, in part, by funding from the J.W. Fulbright Foundation; Office of Naval Research; Center for Neurotechnology Studies of the Potomac Institute for Policy Studies; and the Neuroethics Studies Program of the Center for Clinical Bioethics, Georgetown University Medical Center, Washington, DC, USA (JG). The authors gratefully acknowledge the assistance of Daniel Howlader and Sherry Loveless in the preparation of this manuscript.

References

Bell, C. Neurons for peace: Take the pledge, brain scientists. New Scientist (2010) 205: 24-25.

Benanti, P. From neuroskepticism to neuroethics: Role morality in neuroscience that becomes neurotechnology. American Journal of Bioethics – Neuroscience (2010) 1: 39-40.

Benedikter R., Giordano J., FitzGerald, K. The future of the self-image of the human being in the age of transhumanism, neurotechnology and global transition. Futures: The Journal for Policy, Planning and Futures Studies (2010) 41: 1102–1109

Bitzinger, R.A. Civil-Military Integration and Chinese Military Modernization. Asia-Pacific Center for Security Studies (2004) 3, http://www.apcss.org/Publications/APSSS/Civil-MilitaryIntegration.pdf.

Bower, R., and Giordano, J. The Use of Neuroscience and Neurotechnology in Interrogations: Practical Considerations and Neuroethical Concepts. American Journal of Bioethics – Neuroscience (2012) 3: 3.

Boyd, J.R. The Essence of Winning and Losing, 1995, http://www.danford.net/boyd/essence.htm.

Cacioppo, J.T., Visser, P.S., and Pickett, C.L. *Social neuroscience: People thinking about thinking people*, MIT Press, Cambridge, MA, 2006.

Chalmers, D. Facing Up to the Problem of Consciousness. Journal of Consciousness Studies (1995) 2: 200-219.

Crawford, M.B. The limits of neuro-talk. In Giordano, J., and Gordijn, B. (eds.). *Scientific and Philosophical Perspectives in Neuroethics*, Cambridge University Press: Cambridge, 2010, pp. 355-369.

Davies, P.H.J. Intelligence and the Machinery of Government: Conceptualizing the Intelligence Community. Public Policy and Administration (2010) 25: 29-46.

Eagleman, D. Incognito: The Secret Lives of the Brain. Pantheon, New York, 2011.

Etzkowitz, H. *The Triple Helix: University-Industry-Government Innovation in Action*, Routledge, New York, 2008.

Evans-Pritchard, A. France cuts frantically as Italy nears debt spiral. The Telegraph, November 8, 2011, http://www.telegraph.co.uk/finance/financialcrisis/8875444/France-cuts-frantically-as-Italy-nears-debt-spiral.html.

Fields, J. Capital Punishment: Organic Basis of Criminal Behavior. Hopkins Undergraduate Research Journal (2006) 6: 23-25.

Forsythe C., and Giordano, J. On the need for neurotechnology in the national intelligence and defense agenda: Scope and trajectory. Synesis (2011) 2: T5-T8.

Giordano, J. Integrative convergence in neuroscience: trajectories, problems and the need for a progressive neurobioethics. In Vaseashta, A., Braman, E., Sussman, P. (Eds.), *Technological Innovation in Sensing and Detecting Chemical, Biological, Radiological, Nuclear Threats and Ecological Terrorism (NATO Science for Peace and Security Series)*, Springer, New York, 2012a.

Giordano, J. Introduction. In Giordano, J., and Gordijn, B. (eds.). *Scientific and Philosophical Perspectives in Neuroethics*, Cambridge University Press: Cambridge, 2010, pp. xxv-xxx.

Giordano, J. Neuroethics- two interacting traditions as a viable meta-ethics? American Journal of Bioethics – Neuroscience (2011a) 3: 23-25.

Giordano, J. Neuroethics: Traditions, tasks and values. Human Prospect (2011b) 1: 2-8.

Giordano, J. Neurotechnology as Demiurgical Force: Avoiding Icarus' Folly. In Giordano, J. (Ed.), *Neurotechnology: Premises, Potential and Problems*, CRC Press, Boca Raton, Florida, 2012c, pp. 1-14.

Giordano, J. (Ed.) *Neurotechnology: Premises, Potential and Problems*. CRC Press, Boca Raton, Florida, 2012c.

Giordano J., and Benedikter, R. An early - and necessary - flight of the Owl of Minerva: Neuroscience, neurotechnology, human socio-cultural boundaries, and the importance of neuroethics. Journal of Evolution and Technology (2012a) 22: 14-25.

Giordano J., and Benedikter, R. Neurotechnology, Culture, and the Need for a Cosmopolitan Neuroethics. In Giordano, J. (Ed.), *Neurotechnology: Premises, Potential and Problems*, CRC Press, Boca Raton, Florida, 2012b, pp. 233-242.

Giordano, J., Benedikter, R., and Kohls, N.B. Neuroscience and the importance of a neurobioethics: A reflection upon Fritz Jahr. In Muzur, A., Sass H-M. (eds.). *Fritz Jahr and the Foundations of Integrative Bioethics*, LIT Verlag Münster, Berlin, 2012.

Giordano, J., Benedikter, R., and Floes, N. Neuroeconomics. An Emerging Field of Theory and Practice. European Business Review (2012) 7: 45-47.

Giordano, J., and DuRousseau, D. Toward right and good use of brain-machine interfacing neurotechnologies: Ethical issues and implications for guidelines and policy. Cognitive Technology (2011) 15: 5-10.

Giordano, J., Engebretson, J., and Benedikter, R. Pain and culture: Considerations for meaning and context. Cambridge Quarterly Review of Healthcare Ethics (2008) 77: 45-59.

Giordano, J., Forsythe, C., and Olds, J. Neuroscience, Neurotechnology, and National Security: The Need for Preparedness and an Ethics of Responsible Action. American Journal of Bioethics – Neuroscience (2010) 1: 35-36.

Giordano J., and Gordijn, B. (Eds.) *Scientific and Philosophical Perspectives in Neuroethics*, Cambridge University Press, Cambridge, 2010.

Glannon, W. *Defining Right and Wrong in Brain Science: Essential Readings in Neuroethics*, Dana Press, New York, 2007.

Gregg, K. Compliance with the Biological and Toxin Weapons Convention (BWC). Biosecurity, FAS in Nutshell, 2010, http://www.fas.org/blog/nutshell/2010/08/compliance-with-the-biological-and-toxin-weapons-convention-bwc/.

Groarke, L. When Two Wrongs Make a Right. Informal Logic Newsletter (1982) 5: 10-13.

Gross, E. Thwarting Terrorist Acts by Attacking the Perpetrators or their Commanders as an Act of Self-defense: Human Rights versus the State's Duty to Protect its Citizens. Temple International & Comparative Law Journal (2001) 15: 196-246.

Illes, J. (Ed.). *Neuroethics*. Oxford University Press, Oxford, 2005.

Jahr, F. Bio-Ethik: eine Umschau uber die ethischen Beziehungen des Menschen zu Tier und Pflanze. Kosmos (1927): 2-4.

Jeannotte, A.M., Schiller, K.N., Reeves, L.M., Derenzo, E.G., and McBride, D.K. Neurotechnology as a public good. In Giordano, J., and Gordijn, B. (eds.). *Scientific and Philosophical Perspectives in Neuroethics*, Cambridge University Press: Cambridge, 2010, pp. 302-320.

Juarrero, A. *Dynamics in Action: Intentional Behavior As a Complex System*, MIT Press, Cambridge, MA, 2002.

Levy, N. Neuroethics: A New Way of Doing Ethics. American Journal of Bioethics – Neuroscience (2011) 2: 3-9.

Merriam-Webster's Collegiate Dictionary. Ecology. In *Merriam-Webster's Collegiate Dictionary*, 11th ed. Merriam-Webster Inc, Springfield, MA, 2004, pp. 394.

Montesquieu, Charles de Secondat. *The Spirit of Laws: A Compendium of the First English Edition*, University of California Press, Berkley, 1978.

National Research Council. *Emerging Cognitive Neuroscience and Related Technologies*, National Academies Press, Washington, DC, 2008.

Orofino, S. Daubert v. Merrell Dow Pharmaceuticals, Inc.: The Battle Over Admissibility Standards for Scientific Evidence in Court. Harvard Journal of Undergraduate Science (1996) 3: 109-111.

Racine E. *Pragmatic Neuroethics,* MIT Press, Cambridge MA, 2010.

Ridley, M. *Nature via nurture: genes, experience and what makes us human*, Harper Collins, London, 2003.

Rees, D., and Rose, S. (Eds.) *The New Brain Sciences: Perils and Prospects*, Cambridge University Press, Cambridge, 2004.

Rose, S. *The Future of the Brain: Promise and Perils of Tomorrow's Neuroscience*, Oxford University Press, Oxford, 2005.

Roskies, A. Neuroethics for the New Millenium. Neuron (2002) 35: 21-23.

Schoner, G., and J.A.S. Kelso. Dynamic pattern generation in behavioral and neural Systems. Science (1985) 239: 1513-1520.

Shaneyfelt, W.L., and Peercy, D.E. A Surety Engineering Framework and Process to Address Ethical, Legal, and Social Issues for Neurotechnologies. In Giordano, J. (Ed.), *Neurotechnology: Premises, Potential and Problems*, CRC Press, Boca Raton, Florida, 2012, pp. 213-232.

Shook, J. and Giordano, J. Toward a principled and cosmopolitan neuroethics to guide neuroscience and neurotechnology from bench to bedside and beyond. Philosophy, Ethics and Humanities in Medicine, (2013; in press)

Swaab, D.F. Developments in neuroscience. In Giordano, J., and Gordijn, B. (eds.). *Scientific and Philosophical Perspectives in Neuroethics*, Cambridge University Press: Cambridge, 2010, pp. 1-36.

VanMeter, J.W. Neuroimaging: Thinking in Pictures. In Giordano, J., and Gordijn, B. (eds.). *Scientific and Philosophical Perspectives in Neuroethics*, Cambridge University Press: Cambridge, 2010, pp. 230-2430.
Vaseashta, A. The Potential Utility of Advanced Sciences Convergence: Analytical Methods to Depict, Assess, and Forecast Trends in Neuroscience and Neurotechnological Developments and Uses. In Giordano, J. (Ed.), *Neurotechnology: Premises, Potential and Problems*, CRC Press, Boca Raton, Florida, 2012, pp. 15-36.
Verplaetse, J., DeSchrijver, J., Vanneste, S., and Braeckman, J. *The Moral Brain,* Springer, Berlin, 2009.
von Bertalanffy, L. *General System Theory: Foundations, Development, Applications*, George Braziller, New York, 1968.
von Weizsacker, E.U., Lovins, A.B., and Lovins, L.H. *Factor Four: Doubling Wealth, Halving Resource Use,* Earthscan, London, 1998.
Wilson, E.O. *Consilience: The Unity of Knowledge*, Knopf, New York, 1998.
Wurzman, R. Inter-disciplinarity and constructs for STEM education: At the edge of the rabbit hole. Synesis 1 (2010): G32-G35.
Wurzman, R., and Giordano, J. Differential susceptibility to plasticity: a 'missing link' between gene-culture co-evolution and neuropsychiatric spectrum disorders?. BMC Medicine (201) 10: 37.
Wurzman, R., and Giordano, J. Neurotechnologies as weapons in national intelligence and defense – An overview. Synesis (2011) 2: T55-T71.

Special Editorial Chapter: An Integrated Approach to Understanding Human Behavior

Lieutenant General Robert E. Schmidle, Jr.
Deputy Commandant for Aviation
United States Marine Corps

The potential impact of recent discoveries in neuroscience on our understanding of human behavior is undeniable. However we should keep in mind that biology is only part of what makes up our human selves and defines us as persons living in a given society and culture. An integrated approach, one that takes into account the influence of the empirical sciences as well as a social psychological framework gives us the most holistic understanding of human behavior. As we think about the behavior of terrorists in particular we need to come to grips with the factors that not only caused them to become terrorists, but the factors that caused us to label them as such.

People don't become terrorists simply because of a chemical imbalance in their brains; they become terrorists because of choices they made that contributed to the discursive development of their terrorist self. That discursive development occurred during interaction, primarily linguistic, with other terrorists, and not because of the firing of specific synapses. As has been mentioned by others in this collection of papers, context is critically important to our understanding of human behavior. Any attempt to understand terrorist behavior without

accounting for the social and culture context in which the terrorist self develops is at best incomplete and at worst completely wrong.

In our examination of terrorist behavior we must resist the temptation to simply reduce human perceptions and emotions to the location and strength of firing synapses. Instead we first need to develop the conceptual framework within which we conduct our empirical investigations. That framework will identify errors in the way our investigation is being conducted. A conceptual framework includes the relevant cultural and historical context and provides a bridge between what makes sense in that context and what is in fact nonsense. Sense making occurs through the use of language and therefore it is by examining the *use* of words in language that we make sense of the empirical data that come from scientific investigations. Without a conceptual framework we fall prey to assigning to individual parts of a person the attributes that the more accurately define the person as a whole.

In their book *Philosophical Foundations of Neuroscience* M.R. Bennett and P.M.S. Hacker take issue with the ascription of psychological predicates to parts of a person vice the whole entity. " Mereology is the logic of part/whole relations. The neuroscientists' mistake of ascribing to the constituent *parts* of an animal attributes that logically apply only to the *whole* animal we shall call 'the Mereological Fallacy' in neuroscience." (PFN 73) They go on to propose that there is also a mereological principle in neuroscience, which they describe as "The principle that psychological predicates which apply only to human beings (or other animals) as wholes cannot intelligibly be applied to their parts, such as the brain…" (Ibid)

It is important to remember that it is not the brain that feels pain but the *person* that feels pain. For example, imagine a situation where a brain scan such as PET or MRI indicates neurological activity associated with pain. The mereological fallacy in this case would be to attribute to that person's brain the feeling of pain, when in fact it is the person who is feeling pain. In order to ascertain whether that person is feeling pain they would have to express or exhibit pain behavior. In this instance either verbally (yelling out) and/or physically (grimacing). Therefore unless there is a tendency toward those expressions the person being observed is not correctly identified by the word 'pain.' You can imagine a case where a PET scan would indicate activity normally associated with pain and yet the patient would say she is not in pain, the opposite is of course the case as well.

The point here is that 'being in pain' is not simply or truly a measure of brain activity alone, it is more accurately the expression of pain behavior that is recognized as such by the local culture in which one lives. A child learns pain behavior by observing others in pain not from a private sensation or feeling that is hers alone to experience. Since it is not a biological process that is responsible for the development the human self the behavior exhibited by a person cannot simply be the result of neural firings in the brain. Rather, those human behaviors are the result of discursive interaction with other humans in the cultural and moral order in which they live.

As we think about human behavior we should consider the proposition that man is an embodied person whose sense of self can only be understood within the context of his moral actions. The source of those moral actions is not material but immaterial; it is the result of myriad influences ranging from Immanuel Kant's 'will' informed by reason to Ludwig Wittgenstein's' non epistemic 'hinges.' Indeed empirical knowledge, i.e. preparing for the winter because we know the weather will be cold, is an undeniably important factor in human behavior but it alone doesn't account for that behavior. Both kinds of knowledge, scientific, (i.e. empirical) and philosophic, (i.e. conceptual) are necessary for our understanding of human actions. This is because we are driven by our biology to live a pleasurable life and at the same time driven by our moral will to make ourselves deserving of such a life.

The point here and throughout my positing of the need for an integrated approach to human behavior is that there is a necessary tension between the duality of absolutes (material and immaterial). This tension must be maintained since it is the dynamics of that tension that gives meaning to our study of the terrorist as a person. It is in fact, this notion of a dualism in constant tension that is key to highlighting the implications of new discoveries in neuroscience. Especially so if those discoveries are going to help us understand the moral choice a terrorist makes as his terrorist self develops and is subsequently sustained. Those implications are neither simply obvious nor purely causal.

In other words there exists an eternal tension between man as an imperfect being driven by biological impulses and man as a striving being driven by a moral law. For the committed person whether a terrorist or a terrorist hunter, life is a struggle to overcome physical and biological limitations in order to live a life informed by one's duty and in accordance with moral law. In the end there is great value in an approach to examining terrorism that holds in tension the two opposing influences of biology and psychology. It is not one view or the other that is or should be portrayed as sufficient for an understanding of human activities.

Examining the development of a terrorist consciousness and its relationship to actions in her world is not a question of simply choosing between science and morality. On this view, I do not believe that we can derive the moral principles that drive human actions from scientific facts. In other words neuroscience can only tell us what "is" not what "ought" to be. However, it is clear that our human behavior can be affected by chemical and structural changes in the brain, and to deny those factors a place in the behavioral equation limits a complete understanding of terrorism.

This leads us to the fundamental issue of evaluating the contributions of neuroscience to the field of human behavior examined throughout this volume. Specifically the issue is the extent to which biology affects behavior. Implied in that discussion is the question of whether there are moral actions that are simply right or wrong regardless of what any particular person or group believes about those actions in their local context. If we believe that there are universal criteria for correct moral conduct then how do we assess responsibility for a person's actions if they are under the influence of chemical or biological factors that make him "not

himself?" This assessment of responsibility also raises the question to what extent, if any are there limitations of contextualization on universal moral principles and actions.

While we may not ever definitively answer the question of responsibility for the development of a terrorist, an integrated approach that takes into account both biology and psychology will produce the greatest improvement in our understanding of terrorism. An increased understanding of terrorism, as a form of human behavior, will aid in the formation of actions to mitigate the opportunities for people to become terrorists. That in turn will lead us towards potential ways to deal with individuals after they have passed the point of turning back on their journey deeper into the terrorist self.

References

Bennett, M. & Hacker, P. (2003) *Philosophical Foundations of Neuroscience*. Malden, MA. Blackwell Publishing.

Appendix: Lexicon

Definition of "Neurodeterrence" – *DiEuliis and Cabayan*

Neurodeterrence refers to the application and consideration of evolutionary neurobiological underpinnings of cognitive and psychosocial behaviors that are important to deterrence theory in the context of conflict. It refers to the inclusion of, a systems understanding of how individuals or groups behave and make decisions, in the development of deterrence strategies. It refers to inclusion of these neurobiological systems, such as neurobehavioral violence or aggression, as a formative and additional component of the evidence base used in formulating deterrence approaches. It assumes the evolutionary progression of warfare between groups and that deterrence as a concept may be a long learned aspect of human psychology.

Axon: a long and nerve-cell process that conducts impulses away from the cell body and to the next neuron or muscle.

Amygdale: The amygdala is mass of nuclei located deep within the limbic system in the temporal lobe of the brain. It is involved in processing emotions and motivations, particularly those that are related to survival (such as fear, anger and pleasure.) The amygdale is also responsible for certain memory storage in the brain which may affect emotional responses to particular events.

Frontal Cortex (FC): The FC is part of the cerebral cortex in either hemisphere of the brain lying directly behind the forehead; it receives input from all of the body's senses and processing. The FC is also responsible for the brain's ability to create long-term plans, governs emotions, and is involved in creativity and original thinking.

Cognition: the collection of mental processes that includes attention, memory, producing and understanding language, learning, reasoning, problem solving, and decision making.

Dendrite: any of the usually branching protoplasmic processes that conduct impulses toward the body of a nerve cell

Glia: sometimes called **neuroglia**, are non-neuronal cells that maintain homeostasis, form myelin, and provide support and protection for neurons in the brain, and for neurons in other parts of the nervous system such as in the autonomic nervous system.

Limbic System: a system of functionally related neural structures (including the amygdale) in the brain that are involved in emotional behavior.

Neuron: is an electrically excitable brain cell that processes and transmits information to individual target cells through specialized electrical and chemical signals.

Neuroeconomics: An interdisciplinary field that seeks to explain human decision making (i.e. the processing of multiple alternatives and selecting a course of action) in the context of economics and neuroscience. It combines discoveries and research methods from neuroscience, experimental and behavioral economics, and cognitive and social psychology. It can also utilize approaches from theoretical biology, computer science, and mathematics. ***Neuroeconomics studies decision making by using a combination of these varied disciplines, avoiding the shortcomings of any single individual approach; as such it offers a parallel to "neurodeterrence" in similar framework.***

Neuroscience: the study of the anatomy, physiology, biochemistry, molecular biology, and pharmacology of the nervous system

Neurotransmitter: a chemical by which a nerve cell communicates with another nerve cell.