Naturalizing Cruelty

G. Randolph Mayes

California State University Sacramento

6000 J Street
Department of Philosophy
Mendocino Hall 3032
Sacramento, California 95819-6033
(916) 278-6424
mayesgr@csus.edu

Abstract: Cruelty is widely regarded to be a uniquely human trait. This follows from a standard definition of cruelty as involving the deliberate infliction of suffering together with the empirical claim that humans are unique in their ability to attribute suffering (or any mental state) to other creatures. In this paper I argue that this definition is not optimum for the purposes of scientific inquiry. I suggest that its intuitive appeal stems from our abhorrence of cruelty, and our corresponding desire to define cruelty in such a way that it is almost always morally wrong. Scientifically speaking this is an arbitrary condition that inhibits our attempt to study cruelty as a natural phenomenon. I propose a fully naturalized definition of cruelty, one that considerably expands the range of creatures and behaviors that may be conceived as cruel.

Keywords: altruism, animals, cruelty, definition, evolution, folk psychology, intentionality, morality, naturalization, suffering.

Naturalizing Cruelty

1. Introduction

Most people think of cruelty as a uniquely human characteristic, though few see it as a normal or healthy one. The few are notable, however. Nietzsche claimed that "Almost everything we call 'higher culture' is based on the spiritualization of cruelty." Stanley Milgram showed that the approval of an authority figure is sufficient for ordinary human beings to perform and rationalize cruel acts. More recently, Philip Zimbardo's Stanford Prison Experiment provided a purely situational model for the shocking cruelty displayed by American military personnel at Abu Ghraib.

The suggestion that cruelty is a pedestrian human characteristic offends liberal sensibilities. Thomas Hobbes, who conceived no limit on the violence that humans would perpetrate in pursuit of their own interests, still did not regard us as naturally constituted to enjoy the suffering of others: "For, that any man should take pleasure in other men's great harms, without other end of his own, I do not conceive it possible." Even if we view cruelty as something of which humans are uniquely capable, we tend to insist with Hobbes that delight in the suffering of others is an aberration of human nature, not an essential aspect of it.

But the nature and extent of cruelty in the world is ultimately an empirical matter, and correct answers to empirical questions respect neither the limits of our imaginations nor

our moral sensibilities. If we are to understand cruelty in empirical terms we must operate with a concept of cruelty that satisfies the aims of natural science. Specifically, we must operate with a definition that allows us the maximum flexibility to integrate evolutionary, behavioral, and neurological evidence of cruelty into a unified picture of the phenomenon. This aim is not facilitated by assuming a priori either that cruelty is uniquely human or that it is a perversion of human nature.

In this essay I argue that significant empirical work on cruelty will be impeded by a working definition of cruelty that is inappropriately burdened by moral intuitions and pre-Darwinian assumptions about the uniqueness of man. I then propose a broader and more fully naturalized definition that allows us to predicate cruelty of non human animals and normal human beings without absurdity.

2. Problems concerning the meaning of cruelty

In a recent target article for *Behavioral and Brain Sciences* entitled "Cruelty's Rewards: the Gratifications of Perpetrators and Spectators," noted sociologist Victor Nell defines cruelty as follows:

> Cruelty is the deliberate infliction of physical or psychological pain on other living creatures, sometimes indifferently, but often with delight. (Nell 2006, 211)

Nell's working definition corresponds roughly to the one offered by the *Oxford English Dictionary: "*... the disposition to inflict suffering; delight in or indifference to the pain or misery of others…" Nell is clear from the beginning of his article that the definition of cruelty, in conjunction with the claim that humans are the only animals capable of attributing mental states to other creatures, implies that cruelty is a uniquely human trait. The paper goes on to argue on the basis of evolutionary, neurological and literary evidence that "cruelty is a behavioral by-product of predation" that has conferred various survival benefits to its practitioners. Nell takes his evolutionary account of cruelty to at least partially vindicate the nativist view that "cruelty is perpetrated… by manifestly normal and decent people" who have not in any way been socialized or trained to exhibit cruel behavior. (Nell, 248)

Nell's paper stands as an important scientific contribution to this neglected topic. However, confusion about the meaning of cruelty is apparent both within the paper itself, as well within a large number of responses to the paper. Of the 23 scholars and scientists who participated in the open peer commentary to Nell's target article several registered comments and criticisms that bear on the definition of cruelty. The noteworthy ones fall roughly into two categories: (1) claims that are openly skeptical of Nell's definition; (2) claims that implicitly assume a different definition.

2a. Explicit disagreement about the definition of 'cruelty'

Here are some examples of peer comments that are either explicitly critical of Nell's definition, or skeptical of implications of the definition in conjunction with certain empirical claims concerning the cognitive capacities of non human animals.

> Implicit in the definition [of cruelty] is the notion that humans should be a bit more morally advanced than sub-human primates or other animals. Because of his exclusion of … cruelty to behavior solely in humans, [Nell] ignores the fact that the situations which he documents so well of cruelty in mankind have strong parallels in other animals (Dallman 2006, 227).

> By focusing on intent as the basis for defining cruelty, serious forms of animal harm such as hoarding are minimized because the perpetrator lacks clear intent to harm (Herzog and Arluke 2006, 230).

> May we need a better taxonomy of cruelty? Can one have cruelty without the reflective desire to impart suffering? If "intention to inflict pain" is critical for the concept, how can one evaluate and defend knowledge derived from animals? … Is a critical crux of cruelty that animalian … predatory systems generate primal intentions that are integrally linked to aroused action tendencies (Panskepp 2006, 234)?

> My one quibble with Nell concerns [his restriction of] cruelty to hominids, starting with *Homo erectus*…. I would hesitate to deny a priori the capacity for cruelty in intelligent predator species such as orcas. Almost every claim for human behavioral uniqueness has bitten ethological dust (van Den Berghe 2006, 245).

With slightly different emphases all of these remarks express doubts about the scientific merit of an inquiry that defines a phenomenon so narrowly that its existence in more than one species is ruled out from the beginning.

2b. Empirical claims based on a different definition of 'cruelty'

Several other peer comments seem obviously false on Nell's definition of cruelty. Consider the following remark by Ainslie.

> It is not clear whether a cat plays with a mouse partially in order to savor the distress of the victim, or merely since it is an optimally challenging game (Ainslie 2006, 224).

The suggestion that a cat might be interpreted as "savoring the distress of its victim" is clearly intended to imply that cats may be conceived as cruel. But, in conjunction with the empirical assumption that cats are incapable of attributing mental states to their victims, Nell's definition clearly implies that cats are incapable of cruelty. Here are some other comments of this nature.

> Cruelty is evident in the play and interactions of quite small children….Once they can coordinate intentional movements, infants and toddlers show… they are readily capable of inflicting pain on others (Kraemer 2006, 233).

> I suggest that cruelty is linked to coalitional aggression and same species killing, and probably goes back 7 million years or more to a common ancestor with chimpanzees (Potts 2006, 238).

> The "territorial warfare" of chimpanzee males is marked by continued beating, biting, and pounding that do not always end with the death of the victim…. If these are markers of cruelty, then chimpanzees … seem to show them (Schuster 2006, 241).

If we use Nell's definition we are required to read the authors as attributing advanced cognitive capacities to cats, small children, ancestral hominids and chimpanzees. This is clearly uncharitable. Such creatures are presumably capable of responding at some level to the affective states of others, but is highly implausible to suggest that these responses are regulated by a deliberative process, which is what Nell's definition requires. The charitable reading of these remarks is that the authors are implicitly rejecting Nell's definition by using the term in such a way that a broader range of behaviors may be meaningfully described as cruel.

3. A conventionalist response

Regarding both sets of comments above one might respond with a yawn that the issues raised are entirely conventional. From a logical point of view it simply does not matter whether we restrict term "cruelty" to the deliberate infliction of suffering or broaden its scope and then go on to distinguish between deliberate and non deliberate forms of cruelty. This is certainly true. The problem is that this is not a purely logical matter. What productive science requires of linguistic conventions is not only conceptual clarity but theoretical fertility. Whether or not a linguistic convention will prove fertile is not something that we can know *a priori*. However, we can be reasonably assured that

conventions that researchers in the field regard as cumbersome, unintuitive, idiosyncratic, or burdened with unnecessary assumptions will tend to have a prophylactic effect.

We may appreciate the significance of this point by considering a related example: the concept of rape. Although 'rape' is defined in different ways for different purposes, it is ordinarily understood to mean forced sexual copulation. Like cruelty, rape is widely seen as a uniquely human phenomenon. This is partly because rape is a crime, but it is also because we make a distinction between the human and brute uses of force. The brute use of force occurs without any awareness or concern about the will or the interests of the victim; it is simply the force of nature. A uniquely human use of force, on the other hand, may be thought to depend on an assailant's understanding that the victim has desires and interests that are contrary to what the aims of the assailant.

Do nonhuman animals commit rape? Surely this is an open empirical question. But if rape by definition requires a deliberate use of force then our answer is at hand: no. Again, there are no purely logical reasons we must be open to the concept of non deliberate rape. Ethological descriptions of rape-like behavior in a variety of species including spiders, dolphins, chimpanzees and elephants do not suffer a loss of content if we simply speak of them as engaging in (brute) forced copulation rather than rape. Yet when we constrain the use of the term 'rape' so that it cannot logically be applied to non human behavior we encourage the assumption that rape in humans and forcible copulation in non humans are essentially different phenomena that require different

explanations. This assumption can have the effect of vetoing any evolutionary approach to rape, e.g Palmer (1989), as confused or wrong headed.

We should note that it is possible to remain open to a global evolutionary account of the origins of rape, while insisting that the correct proximate explanations of rape itself make essential reference to a rapist's beliefs about the desires of his victim. But even if we were to assume that folk psychological explanations necessarily trump any neurobiological account that makes no reference to intentional states at all, this view is problematic. For example, it rules out the possibility that most human rape behavior results from the *absence* of such beliefs or their pathological failure to disinhibit a rapist from his actions.

So a conventionalist response to the problem of definition simply fails to take into account the fact that definitions of terms do have a causal effect on the direction of research and the types of explanations we are inclined to favor. Since we can not know prior to research which of these will prove most fertile it is reasonable to avoid definitions that are unnecessarily restrictive in the early phases of naturalization.

4. A word about naturalization

Strictly speaking, to naturalize a term is to purge its definition of irreducibly normative, non natural and supernatural concepts. There are many different reasons that one might oppose naturalization, but basically just one reason to support it and that is to remove

obstacles to scientific inquiry. If, for example, we insist on defining human rationality in irreducibly normative terms and we describe rational thought and behavior in terms of the activity of a mental substance that is immune to causal influences, then it will be logically impossible to develop a scientific theory of human rationality. On the other hand if we think of rationality as an evolutionary adaptation and rational thought and behavior in terms of brain activity, then a route to such a theory may exist.

There is quite a bit more to naturalization than ridding the vocabulary of empirically inscrutable terminology, however. Definitions that appear to be fully naturalistic can still have a stifling effect on scientific inquiry. Nell's definition of cruelty, for example, does not explicitly employ irreducibly normative concepts, so in that sense it is already fully naturalized. However, this definition may still be regarded as having been *fixed* by our normative beliefs.

To see what I mean, consider a happier concept that involves many of the same considerations we are now making with respect to cruelty. 'Altruism' is ordinarily defined as unselfish devotion to the welfare of others. According to this definition, a person who receives a benefit from helping another is not acting altruistically. But because it is not clear whether there *are* any altruistic acts on this definition- everyone gets a little psychic income from voluntarily helping another individual- we may suggest modifying it to permit at least some degree of personal benefit. For example we might define an altruistic act as one in which the benefit received is less than the benefit bestowed. Even though the common definition is not irreducibly normative, our reasons

for suggesting this modification would be to bring the term “altruism” more into conformity with a naturalistic view of human motivation. To bring it more in accord with a naturalist view of *non* human motivation, we might also suggest dropping the requirement that altruism be intentional or that it involve the capacity to consider the welfare of another individual.

Because altruism is ordinarily defined so that (in conjunction with standard empirical beliefs about the cognitive capacities of humans vs. non humans) only humans are capable of engaging in altruistic acts, biologists who study apparently self-sacrificing behavior in non human animals currently distinguish between psychological and biological altruism (Wilson, 1991). This latter concept involves no reference to intentionality and it construes ‘welfare’ narrowly in terms of a contribution to reproductive fitness. Although this distinction has proved fertile insofar as it has created a logical space in which to develop an understanding of animal self-sacrifice, it also tends to reinforce the idea that biological altruism is not ‘really’ altruism; that human altruism is something entirely different, requiring a fundamentally different explanatory framework.

It is possible that the human capacity for altruism and cruelty has no connection whatsoever to our evolutionary history; that both violent and self-sacrificing behavior of our non human ancestors bears only a superficial resemblance and has no explanatory significance for real cruelty and altruism. But from a naturalistic point of view this seems very unlikely. Rather, it seems far more likely that our need to insist on a categorical

distinction between brute and human behavior stems from fundamentally moral intuitions: cruelty is always evil; altruism is always good; since animals are not proper subjects of moral evaluation it follows that they are capable of neither.

In general, then, the task of naturalization is not only to produce definitions that make scientific inquiry possible, but to be sure that they serve scientifically plausible assumptions rather than those that originate from and tend to reinforce the idea that certain features of the world are not susceptible to scientific explanation.

5. Naturalizing cruelty

As we have seen, Nell begins his attempt to contribute to an empirical theory of cruelty by accepting something close to the *OED* definition. In response to some of the concerns noted in section 2 above, Nell generally insists that this definition is commensurate with his explanatory aims. In fact, he clarifies the original definition in a way that makes cruelty by non human animals very difficult to imagine. In discussing van den Berghe's and other's suspicions that cruelty is not a uniquely human trait, Nell writes:

> I am not persuaded. The question is whether the intentionality that is evident in these and other examples …. amounts to a theory of mind that enables these demonic males to not only formulate an intention to inflict pain, but to do so *because* that pain would cause the victim to suffer (Nell, 247).

Although it is not clear what Nell means when he speaks of pain causing suffering, his point seems to be this: We may be able to justify attributing a certain rudimentary form of deliberate action to chimpanzees, even the deliberate desire to cause injury to another being. But we have no basis for explaining this behavior as the result of the chimp's intention to provide another being with a certain set of experiences, nor its ability to desire that its own actions should cause the being to have these experiences.

So Nell seems to assume that cruelty can be meaningfully predicated only of a being that (1) possesses a theory of mind (i.e., one that is capable of understanding the behavior of other beings by attributing mental states to them) and (2) whose *own* behavior requires explanation in terms of its disposition to both desire and cause certain types of mental states to occur in other beings.

Additionally, Nell treats the psychological response to a victim's suffering as a criterion. Specifically, while he does not treat deliberate inflictions of suffering as cruel when they are not specifically motivated by the desire to produce suffering (i.e., as with self defense or surgery), he counts *any* deliberate infliction of suffering as cruel if it is followed by pleasure or indifference to that suffering: "Whether or not the conditions for punishment are met, an act is cruel if the perpetrator *or the audience* experiences physiological or psychological arousal triggered by the victim's pain" (Nell, 213, italics added).

While Nell does not actually provide a definition that incorporates all of the properties he claims are essential to cruelty, I suggest that something like the following definition would result from the attempt to do so.

> Cruelty is any deliberate action that (a) is specifically motivated by the desire to produce physical or psychological suffering; or (b) causes one or more beings to take pleasure in or be indifferent to the suffering caused by that action.

This definition captures the following essential features of cruelty according to Nell.

1. Disciplinary punishment is cruel insofar as it necessarily involves the infliction of pain.
2. An action is not cruel, even if the pain is caused deliberately, if is not *specifically motivated* by the desire to cause pain. (In counterfactual terms: all things being equal, the agent would choose to achieve the goal of the action without causing pain if such an option were available.)
3. Pleasure or indifference to perceived suffering is a manifestation of cruelty, even when the action is not specifically motivated to cause suffering.

We will now examine three aspects of this revised definition that seem to be most questionable from a naturalistic point of view. I will call these: (a) the causal condition, (b) the deliberateness condition and (c) the disjunctive condition (of pleasure in *or* indifference to the suffering of others).

5a. The causal condition

The revised definition treats cruelty as a property of actions that *cause* suffering, but this does not fit well with the view that cruelty may also be exemplified in those who are only spectators to the suffering. I think this tension is the result of the fact that we are trying to do justice to two fundamentally different ways of thinking about cruelty. Specifically, cruelty can be predicated of *actions*, as in the statement: It is cruel to tease people. But it can also be predicated of *agents*, as in the statement: Clarissa is cruel.

Ordinary language supports both ways of using the term. However, as Kekes (1996, 237) notes, the attribution of cruelty to actions is best regarded as derivative. "To say that an action is cruel is to say that it is the kind of action that would be performed by a cruel agent…" Although scientific aims permit us to disregard ordinary usage under certain conditions, in this case it is not advisable. Our theoretical starting point is that cruelty is a property with *explanatory* significance; we begin from the assumption that some kinds of actions can be understood as the *result* of the cruelty of the creatures that perform them. A scientific theory of cruelty should attempt to identify the evolutionary history that has resulted in neurobiological and psychological characteristics enabling creatures to act in a particular manner. Indeed, this is precisely what Nell himself aims at.

I suggest that if a scientific theory of cruelty is to break with ordinary usage at this level, it should actually avoid the derivative use of cruelty as a property of actions. Actions that

promote suffering should not be *defined* as cruel, but simply explained, when appropriate, as a result of cruelty.

5b. The deliberateness condition

The requirement that cruelty be deliberate or intentional is susceptible to the points we made above, since it awkwardly constrains the cruelty of spectators to suffering that it is caused by the deliberate action of agents. But the deliberateness condition is problematic for at least two other reasons as well. In discussing these problems I will speak of cruelty primarily as a property of agents, and only derivatively as a property of actions, as I think even Nell's own views find more satisfactory expression on this usage.

First, recall that, according to the revised definition, a cruel action is not just a deliberate action that gives rise to suffering, but rather one whose *purpose* is to cause suffering. To appreciate the significance of this requirement, consider two brothers, Seth and Cyrus, both of whom derive immense pleasure from performing actions that hurt others in any number of ways. The difference between them is this. While Seth immensely enjoys doing things that hurt people, he maximizes his pleasure by *not* reflecting on the suffering of his victims. Because Seth is not a particularly reflective individual, this requirement is something he can easily satisfy. However, it is possible for some people (e.g., his mother) to cause Seth to reflect on the suffering his actions cause, and doing so makes him feel badly for his victims. Cyrus, on the other hand, is always keenly aware of the suffering he is causing, and his pleasure actually *depends* on it. Indeed, because of the

exertion and danger involved in harming people, Cyrus would not be sufficiently motivated to do so if he could not anticipate their actual suffering.

Seth and Cyrus are caricatures, of course, but the properties they possess are just extreme versions of normal human traits. It is the Seth in us whose enjoyment in winning a competition is diminished upon consideration of the pain his competitors experience as a result, something we normally find it easy to ignore. It is our inner Cyrus who is disappointed at our victim's failure to register pain, or when they express something entirely inappropriate, like compassion for or genuine amusement at our sense of accomplishment in having defeated them.

According to Nell's definition, only Cyrus' actions satisfy the conditions of cruelty. However, I think most readers will agree that ordinary usage easily permits us to describe both Seth and Cyrus as cruel individuals. My point here is not that Nell's definition fails to capture ordinary usage, since our strong tendency to let moral considerations inform our ordinary intuitions guarantees that no naturalistic definition will succeed in this aim. The point is that neither ordinary usage *nor* naturalistic considerations require us to draw a sharp line between Seth and Cyrus. When we treat cruelty primarily as a property of agents and only derivatively as a property of actions, then we have no prior reason to think that in developing a scientific theory of cruelty we should *assume* that there is a categorical difference between individuals who both get pleasure from doing harm to others, but whose pleasures can be intensified and diminished in different ways.

The second problem with the deliberateness condition is that it seems to be grounded in something like the principle of double-effect. According to this principle there is an important distinction between the known side effects of an action and its intended consequences. So, for example, if the only way one can see to relieve a person's suffering is to administer what she knows to be a lethal dose of morphine, the principle holds that it is possible to do so with the intention of relieving the suffering, though without the intention to kill. The principle of double effect is sometimes formulated in explicitly normative terms, but even when formulated as a thesis about the nature of human action it is clear that its ultimate purpose is to provide the basis of a moral distinction between behaviorally identical intentional actions with identical known consequences.

The reason our revised definition of cruelty seems to rest on double-effect is that it makes a distinction between an action in which the goal is pain and suffering and one in which "the goal is not to inflict pain, but to cause the adversary's flight, submission, or death….and …pain that is a by-product of treatment intended to cure or heal." (Nell, 213) We preserved this feature of Nell's view in condition (a) of the revised definition, requiring that a cruel action be "specifically motivated" to cause suffering. But what is the basis of this exclusion? Why should we classify only deliberate actions in which suffering is the primary goal as cruel rather than a broader class of deliberate actions in which suffering is not the primary goal, but no less the result of informed deliberate action?

Again, I think that the most plausible answer to this question is that it comports with certain moral intuitions, the same moral intuitions justified by the principle of double effect. Many would agree that an action that is specifically motivated to cause harm is more reprehensible than an action that is known to cause such harm, but which is specifically motivated to bestow some benefit. But a naturalistic theory of cruelty is not built upon moral intuitions. Nell provides no empirical basis for assuming that the neurobiological mechanism that permits us to cause suffering for our pleasure is distinct from the one that allows us to cause suffering for the purpose of repelling an intruder, cleaning a wound, or punishing a child for misbehavior. In fact, we might reasonably suspect that a willingness to cause injury for personal benefit is what initially made it possible to inflict suffering for the benefit of others.

5c. The disjunctive condition

Our revised definition of cruelty subsumes the *OED* requirement that an agent take pleasure in *or* be indifferent to the suffering of its victims. Generally speaking, disjunctive meanings are not conducive to scientific aims. For example, a unified scientific theory of flight would be exacerbated by a definition of "flight" as "a passing through the air *or* through empty space". This definition conforms to ordinary usage but sustained motion through a medium and sustained motion through a vacuum are distinct physical processes.

Nell himself (213) adopts the disjunctive definition, not out of respect for ordinary usage, but because he sees cruelty in terms of "a continuum ranging from instrumental cruelty, marked by the perpetrator's emotional coldness and distance from the victim, to expressive or affective cruelty, marked by the perpetrator's escalating arousal." In fact, there is a psychological principle that links pleasure and indifference, viz, the principle of habituation, which Nell notes here as well: "Cruel acts arouse strong positive or negative emotions in the perpetrator and the audience, although habituation and instrumentality may attenuate them." So it may be that Nell adopts the disjunctive condition, not on intuitive grounds alone, but on the basis of a particular view about the origin of indifference, i.e., that it results from habituation to an initially arousing stimulus.

The problem here is that habituation is just one way in which indifference may occur. My basic indifference to the suffering of millions of sick and starving people in third world countries (relative, say, to my extreme anxiety about the comparatively minor sufferings of my loved ones) is not the result of habituation to an initially arousing stimulus. Human beings simply do not ordinarily have a strong emotional response to the pain or pleasure of people with whom they have (or can imagine) no practical connection. Hence, if indifference to suffering is to be subsumed within a Nell-consistent definition of cruelty it must be narrowly construed as indifference resulting from habituation to an initially pleasurable response. But this effectively means that indifference may be safely removed from the definition, as it is the pleasure that precedes indifference that explains our capacity for cruel behavior.

Since indifference itself has no unique explanatory value, I suggest that the only remaining basis for retaining it in the definition is our moral intuitions. Because both pleasure in and indifference to suffering are morally abhorrent, we naturally assume that they are both essential aspects of cruelty.

6. Desiderata for a definition of cruelty

On the basis of preceding considerations, I suggest that scientific progress in our understanding of cruelty is most likely to be advanced by a definition that is consistent with the following claims.

1. Cruelty should not be regarded primarily as a kind of action. Rather, it should be construed as a behavioral or psychological disposition to be explained in historical and structural terms by evolutionary science and neuroscience respectively.

2. A creature's capacity for cruelty should not be essentially tied to its ability or willingness to cause suffering, since the disposition to enjoy suffering caused by others and the disposition to enjoy suffering caused by the creature itself are both reasonably construed as a manifestation of the same basic trait.

3. Cruelty should not be construed as requiring an agent to grasp the subjective suffering of another living creature, since the pleasure derived from observable signs of injury to

living creatures is the real phenomenon of interest, and this pleasure may sometimes be optimized by the actual failure to reflect on a victim's suffering. This means that cruelty may be legitimately (i.e., either truly or falsely, but not nonsensically) predicated of non human creatures.

4. Cruelty should be defined in terms of psychological dispositions (e.g. pleasure) or neurological pathways (e.g., dopaminergic responses) that reward and reinforce certain kinds of behavior. Specifically, cruelty should not be tied in any essential way to psychological indifference, since indifference has multiple causally unrelated origins and little if any explanatory significance.

7. A naturalized definition of cruelty

In light of the above I propose the following alternative definition of cruelty for scientific purposes.

> Cruelty is a creature's disposition to be rewarded by the perception of injury.

The terms and implications of this definition may be clarified as follows:

1. A *creature* is any living thing. This definition might be expanded to include non living entities like robots depending on future developments in artificial intelligence.
2. A *disposition* is a tendency toward a certain form of behavior. It implies no particular physical mechanism.
3. A *reward* is any external object or event that acts to positively reinforce a creature's behavior. In the case of human cruelty, the perception of injury is a reward which positively reinforces a human's tendency to experience and/or cause or experience injury. A creature may be rewarded by injuries caused by other creatures.
4. The *perception* of injury may range from the actual feeling of pain, to visual, auditory or tactile experience indicating injury to oneself or others. The perception of injury is distinct from injury itself. This implies the following:
    a. The false perception of injury may act as a reward, since the psychological benefit of injuring a creature occurs even if the actual injury does not.
    b. Unperceived injury from which a creature in fact benefits (e.g., the death of another creature competing for a limited food supply) does not activate the creature's reward system and does not affect cruelty.
    c. The perception of injury is cognitively distinct from the reception of injury images or information. Perception of injury implies that the creature's brain has interpreted injury information as a reliable indicator of the external world. The reception of injury information, which may occur in humans through dreams, imagination, and cinematic or literary depictions

of injury, does not imply realistic interpretation. Scientific study of the effects of the latter on the brain reward system must ultimately determine whether the definition of cruelty should be expanded beyond injury perception to include certain non veridical forms of injury image reception.

5. An *injury* is any form of physical or emotional harm to a creature, including oneself. Pain is subsumed as harm, even when no physical damage occurs. This definition might be expanded to include harm done to non sentient beings (like trees or works of art) depending on what neuroscience reveals about the way the brain's reward system responds to harms of this kind.

8. Significance of this definition for research

It should be clear that the proposed definition meets all the criteria specified in section 6. It is, of course, inconsistent with any definition that limits cruelty to a narrower range of behaviors. For example, it is at odds with any definition that explicitly restricts cruelty to (1) human agents, (2) a disposition for the agent to take pleasure in the suffering the agent himself causes, and (3) the enjoyment of serious and unjustified suffering. It is also inconsistent with the modified Nell definition insofar as the latter explicitly restricts cruelty to creatures with mental states like beliefs and intentions and creatures capable of attributing mental states to other beings. The proposed definition also does not follow

Nell in making psychological indifference a criterion of cruelty, though it remains consistent with habituated indifference being an indicator of cruelty.

For the most part, however, the proposed definition is simply more general. Most behavior that would satisfy Nell's definition of cruelty will count as cruel under the proposed definition as well. On the proposed definition the faculties that human beings use to perpetrate and enjoy injury differs from those of other beings mainly insofar as they are more effective for achieving human goals.

As we noted in section 2 above, much peer criticism of Nell's target article is semantic in nature. Our proposed definition would allow Nell's own empirical theory to be expressed in a manner that would focus the scientific debate on empirical and methodological issues. Nell's view that human cruelty is the evolutionary by-product of the rewards of both predation and intraspecific aggression in our non human ancestors ultimately finds clearer expression in a simple distinction between human and non human cruelty and relieves him of the mental gymnastics required to define cruelty in such a way that it plausibly applies to one and only one species.

We noted in section 4 that scientists now routinely distinguish between biological and psychological altruism, and that it is widely assumed that biological altruism, which involves no specific intention to suffer for the benefit of others, is almost certainly an evolutionary precursor to psychological altruism which ordinarily does involve this intention. Using the term 'altruism' to apply to both the behavior of humans and non

humans alike does not in any way obscure the fact that there is a difference between biological and psychological altruism.

By the same token, it can be helpful to distinguish between biological and psychological cruelty. With Nell, we may acknowledge an interesting difference between the ability to cause and be rewarded by the perception of injury, and the ability to cause and be rewarded by the perception of injury as a result of *forming the intention* to cause that injury. Though these are clearly related, the latter obviously requires a great deal more brain power than the former. Hence, it may (currently) be unique to humans, and may therefore be one of the more interesting manifestations of cruelty from the point of view of cultural anthropology, ethics, or jurisprudence. On the other hand, those who set themselves the task of providing a general theory of cruelty should understand this to consist partly in showing how the extraordinary varieties of psychological cruelty (torture, punishment, delayed gratification, negative reinforcement, child abuse, ridicule, sadism, mortification, schadenfreude and reading philosophy) may have emerged from its biological counterpart.

References

Ainslie, George: 2006, What Good are Facts? The "Drug" Value of Money as an Exemplar of All Non-instrumental Value, Behavioral and Brain Science, 29, 176-177.

Dallman, Mary F.: 2006, Make Love, Not War, Behavioral and Brain Sciences, 29, 227-228.

Herzog, Harold and Arluke, Arnold: 2006, Human–Animal Connections: Recent Findings on the Anthrozoology of Cruelty, Behavioral and Brain Sciences, 29, 230-231.

Hobbes, Thomas: 2004, Leviathan, Kessinger Publishing.

Kekes, John: 1996, Cruelty and Liberalism, Ethics 106 834-844.

Kraemer, Sebastian: 2006, The Cruelty of Older Infants and Toddlers, Behavioral and Brain Sciences, 29, 233-234.

Milgram, Stanley: 1974, Obedience to Authority, New York: Harper Collins.

Nell, Victor: 2006, Cruelty's Rewards: The Gratifications of Perpetrators and Spectators, Behavioral and Brain Sciences, 29, 211-224.

Nietzsche, Friedrich: 1966, Beyond Good and Evil, trans. by Walter Kaufmann, New York: Vintage Books.

Palmer, Craig: 1989, Rape in Nonhuman Animal Species: Definitions, Evidence, and Implications, Journal of Sex Research 26, 355-374.

Panksepp, Jaak: 2006, The Affective Neuroeconomics of Social Brains: One Man's Cruelty is Another's Suffering, Behavioral and Brain Sciences, 29, 234-235.

Potts, Malcolm: 2006, Cruelty's Utility: The Evolutions of Same-Species Killing, Behavioral and Brain Sciences, 29, 238.

Rorty, Richard: 1989, Contingency, Irony and Solidarity, Cambridge: Cambridge University Press.

Schuster, Richard: 2006, Nice Idea but Is It Science? Behavioral and Brains Sciences, 29, 240-241.

van den Berghe, Pierre L.: 2006 Cruelty, Age, and Thanatourism, Behavioral and Brain Sciences, 29, 245.

Wilson David: 1991, On the Relationship between Evolutionary and Psychological Definitions of Altruism and Egoism, Biology and Philosophy, 7 61-68.

Zimbardo, Philip: 2007, The Lucifer Effect: Understanding How Good People Turn Evil, New York: Random House.