December 17, 2011

For T. Nadelhoffer (ed.), *The Future of Punishment.* New York: Oxford University Press.

Brute Retributivism[*]

Shaun Nichols
Department of Philosophy
University of Arizona

> Beside good and evil, or in other words, pain and pleasure, the direct passions frequently arise from a natural impulse or instinct, which is perfectly unaccountable. Of this kind is the desire of punishment to our enemies, and of happiness to our friends; hunger, lust, and a few other bodily appetites (Hume, *Treatise* 2.3.9).

Our norms of retributive justice are notoriously difficult to justify. This leads many philosophers to reject the legitimacy of those norms. Other philosophers attempt to justify retributivism by relying on moral realism, arguing that retributivism is one of the moral truths. In this chapter, I will assume that moral realism is false and proceed to argue that the retributive norm is part of a set of norms that do not need justification. That is, it is appropriate to retain the norm even if we cannot provide further reasons in favor of the norm. Of course, this does not mean that retributivism should stand, all things considered. To determine whether we should ultimately preserve our retributive norms will depend on a number of other factors, including both the expected utilities and our other core values. Here the aim is simply to keep retributivism on the table.

1. A bare retributivist norm

Perhaps the most famous proclamation on retributivism comes from Kant's remarks on the last murderer:

> Even if a civil society were to be dissolved by the consent of all its members (e.g., if a people inhabiting an island decided to separate and disperse throughout the world), the last murderer remaining in prison would first have to be executed, so that each has done to him what his deeds deserve and blood guilt does not cling to the people for not having insisted upon this punishment; for otherwise the people can be regarded as collaborators in his public violation of justice. (*Metaphysics of Morals*)

[*] I am grateful to Jerry Gaus, John Doris, Michael Gill, Dominic Murphy, Thomas Nadelhoffer, Tamler Sommers, Mark Timmons, and Manuel Vargas for discussion and comments on earlier drafts of this paper. Thanks also to audiences at St. Louis University and the 2010 meeting of the Society for Applied Philosophy.

This passage succeeds in illustrating a crucial feature of retributivism – it is a backward-looking, non-consequentialist view. But Kant's example also includes several elements that are not essential to retributivism. Most obviously, Kant demands *capital* punishment. Retributivists need not accept this. It further suggests that failing to exact retribution renders one complicit (sharing "blood guilt") in the villainy, but this isn't essential to retributivism either. Kant's passage also presumes that the *state* has the authority to deliver punishment, but this assumption is not built into retributivism (Murphy 1985; Husak 1992; Shafer-Landau 1996).

Even setting aside the additional elements in Kant's exhortation, retributivism has been developed and defined in a number of different ways.[1] The kind of retributivism to be defended here is a basic form. Michael Moore offers a representative statement: "Retributivism is a very straightforward theory of punishment: We are justified in punishing because and only because offenders deserve it." (Moore 1987, 181). According to Anthony Duff, despite the diversity in retributive theories, all attempt to answer the question, "what is the justificatory relationship between crime and punishment that the idea of desert is supposed to capture: why do the guilty 'deserve to suffer'" (Duff 2008). There are two distinct elements in these characterizations of retributivist theories of punishment:

i. A norm that wrongdoers should be punished because (and only because) of their past wrongdoing.
ii. A justification for the norm.

For reasons that are central to the aim of this paper, I want to restrict the retributivist theory to the first factor – a retrospective norm prescribing the punishment of wrongdoers.[2] Call this the *bare retributive norm*.[3] As stated, this is a retrospective, backward-looking norm. This distinguishes the theory from other prominent approaches, all of which are prospective. Most obviously, a consequentialist ethics of punishment looks to future benefits of punishment (e.g. Bentham, Rawls 1955). But humanitarian approaches (e.g. Menninger 1968) and moral education approaches (e.g. Hampton 1984) are equally forward looking. So too are restorative accounts of punishment, which aim to repair relationships (e.g. Braithwaite 1999). By contrast, the bare retributive norm simply says that wrongdoers should be punished for their past actions.[4]

---

[1] See Cottingham (1979) for a cranky paper on the diversity of uses of the term "retributivism" in retributive theories of punishment.

[2] This is still, of course, quite vague. For instance, it leaves open what makes one a 'wrongdoer' in the relevant sense.

[3] Etymologically, "retribution" is tied to the notion of repayment. And this accords with many prominent accounts of retributive punishment, including Kant's. But the notion of payback is itself morally thick. For it loads into the norm a *reason* for punishment ("the guilty should be punished because it's payback"). As a result, I want to avoid identifying retributivism with a pay-back theory of punishment.

[4] Desert is often invoked in characterizing retributive theories of punishment (e.g. Moore 1987, 181; Duff 2008). The retributivist typically maintains that wrongdoers *deserve* to be punished for their past wrongs. In some cases, saying "wrongdoers deserve to be punished" might just be a restatement of the bare retributive norm. But often the appeal to desert seems to aspire to provide a substantive *justification* for retributivism – the (alleged) fact that wrongdoers deserve to be punished is supposed to be a value-adding reason for endorsing the bare retributive norm. Again, I want to focus narrowly on the bare retributive norm, not on the justifications for it. As a result, insofar as the appeal to desert is supposed to provide a deeper justification for retributive

2. The bare retributive norm and the folk

It is widely assumed that ordinary people are retributivists. Indeed, it's often taken for granted that retributivism has its roots in ordinary thought (e.g., Hume 1739/1964 section 2.3.9, Smith 1790/1982, pp. 77-8; pp. 87-91, Mackie 1982). But it is an empirical claim, so it's worth looking to some data. Is there, in folk ethics, a bare retributive norm?

Perhaps the most celebrated empirical work on punishment in recent years comes from experimental economics.[5] Using several different economic games, researchers have shown that participants will punish – deduct money from – those who are perceived to have acted unfairly.[6] In one study, groups of four participants play public goods games anonymously on computers. Each participant gets an allotment of money and is allowed to use the money to invest in a common fund. For each 1 monetary unit an individual invests, each of the four players (including the investor) gets .4 units. This provides a net benefit for the group, but the investor himself loses on the transaction. Participants play a series of such games, never with any of the same players. After each game, participants are informed about the contributions of each player and are given a chance to pay to have money deducted from any of the other players in the game. For every 1 monetary unit the punisher pays, 3 monetary units are deducted from the punishee's fund. The participants know that they will not play another game with any of these particular players, so punishing apparently has no future benefit for the punisher. Nonetheless, participants often paid to punish those who contributed less than average (Fehr & Gächter 2002, p. 137).

In a striking extension of this work, Fehr & Fischbacher (2004) explored whether external observers would punish players in an economic game. The third party observed two participants engage in a prisoner's dilemma game. The third party was then given money and asked whether he would use some of this money to pay to deduct funds from either player. Almost half opted to punish a defector in a scenario in which the other player had cooperated (73).

In both of these studies, the motivation for punishment does not seem to be anything like explicit considerations about utilities – the punisher and punishee both lose money, and the punishers have little reason to think the punishment will materially improve the situation of the

---

punishment, I want to set aside the notion of desert.

[5] There is a complementary line of research in social psychology on people's judgments about sentencing of criminals. This work indicates that people are largely retributivists about sentencing (see especially Carlsmith et al. 2002 and Carlsmith 2008). For the purposes of this paper, I want to focus on how people think about punishment in interpersonal interactions rather than how they think about institutionalized forms of criminal punishment. The work in experimental economics is especially apt for this. These experiments involve interaction between individuals in novel scenarios. As a result, there is less chance that subjects are relying on social scripts about criminal punishment. In the economic experiments there is no suggestion of criminal behavior, and both the (apparent) wrongdoing and the punishment are fairly minimal. The experiments even avoid explicit mention of "wrongful" action or "punishment".

[6] I follow experimental economists in categorizing these actions as punishment. This categorization is perfectly appropriate on various philosophical definitions of punishment that aren't narrowly focused on state-sponsored punishment (see e.g., Baier 1955, Benn 1967, Gaus 2011).

other players. As a result, Fehr and colleagues dub it "altruistic" punishment.

De Quervain and colleagues used brain imaging techniques to explore the neural underpinning of this kind of punishment. For their experiment, participants played the trust game (de Quervain 2004). The basic framework of the trust game is explained to all players before the game begins. The trust game is played anonymously by two people. One player, A, is given a sum of money and given the option to send the money to B. If A does send the money the amount will be quadrupled. Then B is given the opportunity to send A half of the money or none of it. In the de Quervain experiment, after playing the trust game, A is given the option to deduct money from B. As expected, when B did not send back any money, A often opted to deduct money from B. And when A did engage in such punishment, there was increased activity in brain regions known to be associated with reward (viz., caudate nucleus and dorsal striatum). Thus, it seems that punishing, at least in these games, is rewarding.

These are beautiful and justly famous results, but they don't show that people have a retributive norm. There are many activities that I expect myself to do and that no doubt stimulate my reward center but that I don't think that I *should* do. For instance, I expect that I will eat bacon later, and the reward structures in my brain will likely respond enthusiastically. But do I think I *should* eat bacon? Do I think that it is a good thing to eat bacon? Of course not. On neither the moral nor the prudential level do I think I should eat bacon. But that will not prevent my reward structures from firing when I eat bacon. Indeed, as de Quervain and colleagues are happy to note, the structures activated by punishment are also activated by consumption of cocaine and nicotine (Breiter et al. 1997; Stein et al. 1998). So, while the imaging results indicate that punishment is *yummy*, they don't show that punishment is guided by a retributive norm. The results don't speak to whether people normatively endorse the behavior that they find rewarding.

Fortunately, it's pretty easy to investigate whether or not people endorse this kind of punitive behavior. We can just *ask* them.[7] I conducted a small study to investigate the matter.[8] Participants were presented with one of three versions of the trust game. In all cases, participants were told about an instance of the trust game in which A sends his money to B, and B sends no money back. And in all cases, participants were asked whether an agent *should* deduct money from B. In version 1, A has to pay to deduct money from B. In that case, roughly half of the participants said that A should pay to deduct money. This is similar to the proportions found in many of the one-shot pay-to-punish experiments (e.g. Fehr & Fischbacher 2004). In version 2, everything was the same, but A can deduct money from B without A himself having to pay anything.[9] In that case, nearly all of the participants said that A should deduct money from B. In the final version, a third party, C, observes the entire set of transactions between A and B. Then C has an opportunity to deduct money from B without C having to pay anything. Nearly all of the participants said that C should deduct money from B.[10]

---

[7] Experimental economists tend not to be very interested in people's explanations for their actions. This might be part of the reason that the topic isn't explored in the original studies.

[8] Participants were recruited through Amazon's Mechanical Turk (https://requester.mturk.com/mturk/welcome). Users of the site can fill out surveys for modest compensation. Buhrmeister et al. (2011) provide evidence that the data obtained through this site is of comparable quality to that obtained in standard psychology subject pools.

[9] de Quervain and colleagues call this version "intentional and free" (1256).

[10] The responses were significantly greater than what would be expected by chance alone in both

These results indicate that people do indeed endorse the punitive behavior that we see in the experimental economics games, particularly when it isn't costly. So, the case of punishment is importantly different from my bacon eating. While I don't normatively endorse my bacon consumption, people do normatively endorse punishment. There is a further question however. Just because there is normative endorsement of the punishment behavior doesn't yet mean that the norm is retributive. To investigate this, we can look at the explanations people gave for their answers. Recall that the critical feature of the bare retributive norm is that it is distinctly backward looking. All of the prominent alternatives, by contrast, are forward looking. When we examine the explanations participants gave for saying that money should be deducted from B, some of the explanations were, in fact, forwarding looking. Here are a couple of examples:

> "B was being selfish and should have some money deducted from him so he will learn no to be so self centered."
>
> "I think that A should subtract the money because B should realize that with reward comes responsibility."

Although a few of the explanations included forward looking considerations, the vast bulk of explanations make no reference at all to the future, focusing instead on *what B did.* Participants were much more likely to invoke only backward looking factors in their explanations.[11] Here are some representative examples:

> "It was a selfish act for B not to send any money back to A"
>
> "B made money off of A's contribution but A did not get anything in return. B should be punished."
>
> "It is unfair that A gave B $12 and B didn't send anything back to A"
>
> "B is a jerk and didn't give anything back."

These explanations share the idea that money should be deducted from B *because of what B did*.

The fact that most subjects don't mention forward-looking factors doesn't exclude the possibility that consequentialist considerations played a critical causal role in their judgments. Indeed, one natural possibility is that people punish other players because of the communicative effects of such punishment (e.g., for moral education or norm reinforcement). After all, in these experiments, the players expect the punishee to be well aware of having their welfare reduced. To address this limitation of extant work, a recent study explored punishment in an economic game in which it was quite explicit that the punishee would be unaware of any welfare reduction.

---

condition 2 (χ^2 goodness-of-fit (1, N=15) = 11.267, p<.001, two-tailed) and condition 3 (χ^2 goodness-of-fit (1, N=15) = 11.267, p<.001, two-tailed). To provide a contrast, a fourth condition was run in which a different economic game was described, modeled on an egalitarian-motive game (cf. Dawes et al. 2007). In this game, a computer randomly assigned money to two players. Sometimes the assignments were equal and sometimes they were disproportionate. Participants were asked to judge whether A should deduct money from B if the computer randomly gave B all of the money and A none of it. In this scenario, most people (12 out of 15) said that A *shouldn't* deduct money from B. This was significantly different from chance (χ^2 goodness-of-fit (1, N=15) = 5.4, p<.05), and much different from the parallel trust game (condition 2) (χ^2 (2, N=30) = 16.43, p<.0001, two-tailed).

[11] Explanations were categorized by independent coders (agreement was over 95%). Participants gave backward looking explanations significantly more frequently than would be expected by chance (χ^2 goodness-of-fit (1, N=35) = 17.857, p < .0001).

Even in this case, however, people still preferred to punish players who behave in ostensibly unfair ways (Nadelhoffer et al. forthcoming). Given that the punishees would not even be aware of the punishment, it is difficult to see how consequentialist reasoning could be driving subjects preferences. Rather, such punishment is plausibly driven by a retributive norm that operates without consulting consequentialist considerations.[12] This non-consequentialism is not a unique feature of the retributive norm. Many of our ethical norms operate independently of consequentialist reasoning (Gill & Nichols 2008). Consider the norm prohibiting incest. If subjects are presented with a vignette in which a brother and sister have consensual sex, with all the prophylactic caution in the world, many subjects persist in thinking that the action was wrong, even though they can't justify their judgment (Haidt et al. 2000). Presumably what is going on here is that subjects have a norm prohibiting incest, and the behavior is categorized as an instance of the prohibited kind. Critically, however, people embrace this norm without having a deeper justification for it. The incest norm is *inferentially basic* – it is not the product of consciously available inferences from other norms or facts. Inferentially basic norms like the incest norm are at the bottom of the normative pile. The retributive norm is a member of this special class of norms. If the bare retributive norm and its downstream consequences were extirpated from our psychology, the norm would not just regenerate from our other consciously available forward-looking norms and values. The bare retributive norm is a basic, independent part of our moral worldview.

3. Retributivism and the emotions

It's widely assumed that retributivism itself is a product of rather base emotional reactions like anger or resentment. Anger is important to retributivism, but the relation between anger and the retributive norm is indirect. Let's start by looking at evidence on anger and retribution in our comfortable domain of economic games.

In the *ultimatum* game –the most intensively studied economic game – one of two anonymous players is randomly assigned to be the *proposer*. The proposer is given a sum of money and told to offer a division of the money to the other player, the *responder*. If the responder accepts the offer, both players get the money, but if the responder rejects the offer, neither gets any of the money. A large body of evidence indicates that when the proposer makes a highly inequitable offer (80/20 or 90/10), responders often refuse the offer, turning down free money.[13] What precipitates this behavior? Anger, apparently. In an early study on the matter, Pillutla and Murninghan (1996) asked participants in an ultimatum game two open-ended questions: "How did you react when you received your offer?" and "How did you feel?" (215). The responses were coded for reports of anger and perceived unfairness. Subsequent analyses showed that when subjects mentioned anger in their responses, this was an excellent predictor of whether they had rejected the offer, even better than perceived unfairness. The researchers offer a simple explanation: when participants regard the offer as unfair, this often triggers anger, and anger increases the tendency to reject the offer (220; see also Bosman & van Winden 2002, p.

[12] Indeed, people tend to agree with a quite explicitly non-consequentialist justification of punishment: "People who commit crimes deserve to be punished even if punishing them won't produce any positive benefits to either the offender or society—e.g., rehabilitation, deterring other would-be offenders, etc." (Nadelhoffer et al. forthcoming).
[13] It's unclear whether we should count these behaviors as instances of punishment (see Gaus 2011). Experimental economists often do presume that rejecting an offer amounts to punishment (e.g., Bolton & Zwick 1995; Sanfey et al. 2003). We don't need to take a stand on this here.

159; Hopfensitz & Reuben 2009).

Fehr and Gächter offer a similar explanation for why people pay to punish in public goods games. They asked their participants to imagine accidently encountering a fellow player who had invested much less than everyone else in the group. Participants were asked to indicate their feeling toward this person (Fehr & Gächter 2002, 139). As expected, participants indicated that they would feel high levels of anger towards this person. Fehr and Gächter draw on this finding, in conjunction with the pattern of punishment behavior, to argue that anger (and perhaps related emotions) plays a critical role in generating punishment in these games.

Anger plausibly plays an on-line role in motivating punishment in economic games. This fits well with a standard picture of the *function* of anger, which dates back at least to Darwin. According to Darwin, anger serves to motivate retaliation. He writes:

> animals of all kinds, and their progenitors before them, when attacked or threatened by an enemy, have exerted their utmost powers in fighting and in defending themselves. Unless an animal does thus act, or has the intention, or at least the desire, to attack its enemy, it cannot properly be said to be enraged (1872, 74).

Although punishment behavior is plausibly driven by anger, it would be a serious error to conclude that retributive judgment is identical to feeling anger. I can judge that a given wrongdoer should be punished even if I don't feel any anger, perhaps because the description of the incident is too vague to induce emotion. Similarly, I can think that a class of individuals (say, shoplifters) should be punished while on a perfectly even emotional keel. In addition to the emotion, we also have the retributive *norm.* And the retributive norm can be activated even in the absence of emotional reaction. The reverse is also possible. Anger is sometimes triggered when there is no sensible target of punishment, as when a turn in weather destroys your grand plans for a dayhike. In this case, the emotion is present, but the norm fails to be activated in any familiar way.

There are thus two elements that are important to retributive punishment – anger and a retributive norm. I've stressed their independence, but there is also, I suspect, an important causal connection between the two. The retributive norm would have inherited cultural strength from its natural links to anger. The idea here draws on epidemiological approaches to cultural evolution (Sperber 1996, Boyer 2000, Nichols 2002). By identifying characteristic features of human psychology, we can get some idea about which kinds of cultural items will be attractive to creatures like us. Many different kinds of emotions – e.g., anger, fear, jealousy, disgust, and sympathy – are characteristic features of human psychology. These emotions make some things attractive and others aversive. As a result, emotions plausibly play an important role in influencing which cultural items are likely to persist. Norms and other cultural items would have increased attractiveness when they resonate with common emotional endowments. For example, etiquette norms prohibiting the display of bodily fluids seem to be preserved once they are introduced into the culture, and a plausible explanation for this is that these prohibitions resonate with our natural proclivity to feel disgust at bodily fluids (Nichols 2004). In the present context, the relation between emotion, motivation, and norms is rather different, because the retributive norm is prescriptive rather than proscriptive. But the following is a plausible principle concerning the cultural evolution of prescriptive norms:

> *Ceteris paribus,* prescriptive norms that encourage behavior that we are naturally motivated to perform will enjoy an advantage over prescriptive norms that lack any such

connection to motivation.[14]

This principle would apply to the case of retributive norms. For the norm that *wrongdoers should be punished* resonates with our natural anger-driven motivation to retaliate against (perceived) wrongdoers. We *want* to retaliate against wrongdoers. This anger-driven motivation would plausibly contribute to the cultural heft of a norm that prescribes inflicting harm on wrongdoers.

Although anger undergirds the retributive norm, anger itself is unruly. In many cultures, including our own recent past, anger-driven retaliation manifested in the practice of blood feud. Blood feuds are not characterized by carefully measured responses. In retaliating against an affront, the aim is often to demonstrate one's power, not carefully measure the harm and meet it in proportional kind. Thus, retaliation in blood feud often involves harming uninvolved family members of the enemy (e.g. Miller 1990, 180; 197-8; Sommers forthcoming).[15] This is *not* the narrow retributive norm that we embrace today. Unlike blood feud, the bare retributive norm directly targets the wrongdoer.

Somehow the unruly retaliatory practices associated with blood feud get displaced by the narrow norm of retribution, tied to the wrong-doer and proportionality. As Biblical scholars have observed for decades (see, e.g., Berlin & Brettler 2004), the injunction to take an *eye for an eye* (Ex. 21:23, 24; Lev. 24:19, 20; and Deut. 19:21) was introduced as a more *moderate* approach to retaliation than the dangerously escalating blood feuds (Gen 4: 23-4). We get a similar phenomenon in the history of early English law. As feudalism replaces tribalism in England, compensation takes the place of the feud (Jeffrey 1957, 665; Harding 1966, 21). To pacify the victim or victim's family, the law indicated that the victim needed to be paid. The laws themselves are graphically specific:

> If an ear be struck off, let bot [monetary restitution] be made with 12 shillings.
> If the other ear hear not, let bot be made with 25 shillings.
> If an ear be pierced, let bot be made with 3 shillings.
> If an ear be mutilated, let bot be made with 6 shillings. (Laws of Ethelbert)

Eventually, of course, this system of victim compensation gives way to the idea that striking off an ear is a crime against the state, requiring something more from the wrongdoer than payment of damages (cf. Maine 1861, chap. 10). This development in the history of law is a further

---

[14] In previous work (Nichols 2002, 2004) I formulate an "affective resonance hypothesis" that is limited to proscriptive norms. Here I am extending the idea to prescriptive norms. But there is a complicating feature about prescriptive norms. Insofar as the behavior is motivationally attractive and there are no countervailing considerations, it's not clear that norms play any significant role. Norms become more obviously significant when there are considerations *against* performing the prescribed action. In the interesting cases, there will likely be such considerations. For instance, we have a prescriptive norm of benevolence, and this competes straightforwardly with self interest. Similarly, the prescriptive norm of punishment often competes with material self interest in the short term (though perhaps not in the long term [see Frank 1988]).

[15] Killing the relatives of wrongdoers is even prescribed in an item in the code of Hammurabi. The son of a carpenter might be executed for his father's shoddy work:

> If a builder build a house for some one, and does not construct it properly, and the house which he built fall in and kill …. the son of the owner the son of that builder shall be put to death. (229 & 230).

departure from the anger-driven retaliatory practices of our ancestors.

All of this suggests that our (narrow) retributive norm was not fashioned *ex nihilo*, spewing forth from rationality. Instead, our retributive norm is a product of cultural pruning. The unfocused retaliatory norms and practices of our ancestors were reshaped and refined, leaving us with the vestige we have today. But *anger* was likely a sustaining factor throughout this cultural evolution of punishment norms. Had our ancestors lacked the propensity for anger at wrongdoers, we today would likely not have the retributive norm we do.

4. Justification

Thus far, I've argued that (i) judgments about punishment are guided by a bare retributive norm that is inferentially basic and (ii) anger contributed to the cultural success of this norm. If this psychological-historical account of the bare retributive norm is right, it should not be surprising if we can't justify the retributive norm. We inherited our retributive norm from emotion-driven cultural evolution and not through rational discovery.

Retributivism is, as a matter of fact, notoriously difficult to justify. The temptation is to defend retributivism by pointing to its future benefits: If we are retributive, then this will provide a deterrent. Or, if we are retributive, then this will enhance cooperation. Or, if we are retributive, this will forestall the pursuit of vengeance by the victim. These kinds of proposals are philosophically appealing. But they are intuitively unsatisfying; for they invoke *prospective* reasons for punishing, while the norm is expressly backwards looking (cf. Mackie 1982, 4; Bedau 1978, 616). Insofar as the retributive norm is inferentially basic, it is perhaps unsurprising that these "deeper" justifications are unsatisfying. Consider the inferentially basic norm that parents have special obligations to their children. Attempts to give a "deeper" justification for this norm, e.g., in terms of societal benefits, are bound to seem wrong-headed. So too, the appeal to future benefits is bound to seem intuitively inadequate as a complete justification for the backward-looking retributive norm.

In the face of this, a typical response is to renounce retributivism. Many intellectuals find bare retributivism offensive – it counsels harming someone without getting any offsetting benefit. In this light, people seem to prefer *anything else* – humanitarianism, restorative justice, utilitarianism... Duff makes this point nicely: "Many people, including those who do not take a consequentialist view of other matters, think that any adequate justification of punishment must be basically consequentialist. For we have here a practice which inflicts, indeed seeks to inflict significant hardship or pain: how else could we hope to justify it than by showing that it brings consequential benefits sufficiently large to outweigh, and thus justify, that hardship and pain" (2008).

There is one prominent metaethical view that offers a haven for philosophers trying to defend an uncompromising retributivism: moral realism. According to moral realism, the basic moral truths are "stance-independent"; they are not made true because of our (or anyone else's) attitudes, norms, emotions, etc. (see, e.g., Shafer-Landau 2003, 15). A retributivist can draw on this metaethical view to maintain that the retributive norm reflects a moral truth about the rightness of punishment, a truth that is independent of what our norms or emotions or evaluative reactions happen to be. Michael Moore is perhaps the most prominent advocate of this view in the recent literature. He maintains that retributivism is a moral truth, and he argues that the emotions play an important role in getting us to appreciate this truth: "The emotions are… heuristic guides for us, an extra source of insight into moral truths beyond the knowledge we can

gain from sensory and inferential capacities alone" (Moore 1987, 201; also 189). The moral truth of retributivism is independent of the emotions, but the emotions help us tap into moral truths (1987, 186-7). They are "important but not essential in our reaching moral truths" (202). In the case of retributivism, Moore suggests that the emotions of guilt and fellow feeling lead us to appreciate the truth of retributivism (209ff.).

There are reasons to be skeptical of Moore's appeal to emotions as indicators of moral truth, but I want to focus on a broader point. The retributivist realist is true to the cause – no consequentialist in retributive clothing here. But the attempted defense of retributivism is precarious. For should moral realism be false, retributivist realism leave us with no basis for preserving the bare retributive norm. And moral realism is disputed by a great number of philosophers, starting perhaps with Hutcheson and Hume but continuing in a frenzy of late (e.g., Blackburn 1998, Gibbard 1992, Greene 2008, Harman 1996, Joyce 2002, Mackie 1977, Nichols 2004, Prinz 2007, Sinnott-Armstrong 2006, Street 2006, Timmons 1999). Should this wave of dissent to realism be right, the retributivist seems to be left without a hope of sustaining the view.

5. Ethical conservatism

The tempting defenses of retributivism seem to either fail to be intuitively retributive or they rely on highly contentious metaethical assumptions. I am going to assume that moral realism is false – that there is no moral truth that stands independent of our attitudes and feelings. I will also assume the falsity of other "moral objectivist" views according to which there is a single true morality (e.g. Clarke 1728). With that (major) assumption in place, I suggest that we don't *need* to have a justification for the retributive norm in order for it to retain its legitimacy for us.[16]

In realist domains if we show that a belief in the domain is formed by a process that is neither reliable nor rational, this undercuts any basis for sustaining the belief. This kind of debunking argument is nicely illustrated in Freud's views on religion. Freud said that the reason we believe in God is because of wishful thinking. And wishful thinking is an epistemically terrible basis for belief formation. Hence, our belief in God is in bad epistemic repair – we are not warranted in our belief. Similar kinds of arguments have been suggested for claims in metaphysics of identity (Scholl 2007), metaethics (e.g. Nichols forthcoming), and consciousness (e.g. Fiala et al. forthcoming).

Joshua Greene applies this kind of debunking argument to normative ethics, and in particular, to retributivism (Greene 2008; see also Singer 2005). First, Greene notes that our retributive inclinations are grounded in arational emotions. Attempts at justifying retribution are, Greene suggests, "just rationalizations for our retributivist feelings… the natural history of our retributivist dispositions makes it unlikely that they reflect any sort of deep moral truth" (2008, 71). Greene uses these considerations about the emotional origins as the basis for debunking retributivism and other deontological principles (Greene 2008; see also Singer 2005, 350).

Discovering that arational emotions generate beliefs in metaphysics, metaethics, or religion might be reason to suspend those beliefs. But in domains in which realism is rejected, it is far from clear that the normal justificatory demands on belief formation apply.[17] The vast bulk

[16] At least on some readings, similar views are suggested in Strawson (1962) and Mackie (1982).
[17] Error theory introduces a complication. According to error theorists, moral beliefs presuppose moral objectivism, and since moral objectivism is false, it follows that all moral beliefs are false (Mackie 1977; Joyce 2002). Error theory depends on controversial semantic assumptions, but

of our ordinary ethical worldview likely derives from fundamentally arational emotional processes (Blair 1995, Gill & Nichols 2008, Prinz 2007). For instance, were it not for the fact that we find human suffering aversive, we would likely not have the moral revulsion we do at killing. Nor would we feel the moral obligation for helping strangers. I am assuming here, with irrealism, that there is no ultimate rational justification for these norms. They are the norms we happen to have, given the kinds of emotional propensities we happen to have. To limit our ethics to norms that have some ultimate rational justification would leave us with an ethics more barren than almost anyone would be willing to accept.[18]

Instead of reverting to such an emaciated ethics, we might adopt an *ethical conservatism*, according to which certain ethical norms do not lose their normative legitimacy even if the norms do not derive from the kinds of processes that confer justification in realist domains (Nichols, Timmons, & Lopez forthcoming). If *none* of our ethical beliefs has an ultimate justification, then, barring a complete upheaval of commonsense ethics, we are bound to grant normative clout to *some* moral norms that lack any ultimate justification.

A full defense of this ethically conservative position is clearly beyond the scope of this paper, but it is important to note that not *all* norms have this privileged status. It accrues, we suggest, to norms that run deep in our psychology – norms that are *entrenched* (Nichols et al. forthcoming). Entrenched norms in this sense have three features. First, they are widespread in the community. Second, entrenched norms are inferentially basic, i.e., not inferentially dependent on other norms or facts (see section 2).[19] Finally, entrenched norms are rooted in human emotion – they resonate with our natural emotional endowment. The norm "help

---

more importantly for present purposes, even people who explicitly reject moral objectivism still treat the "moral" norms in much the same way (Nichols 2004). If error theory is right, the retributive norm can be regarded as "quasi-moral".

[18] Peter Singer presents a Greene-style argument that deontological intuitions are based in arational emotions and hence lack any rational justification (2005, 347ff.). However, he holds out optimism that moral skepticism can be avoided, suggesting that there are some *rational intuitions*, like "it is a bad thing if a person is killed" (Singer 2005, 350-351). Singer doesn't actually provide an argument that this intuition is somehow rationally grounded, and, for my part, I find it hard to see why we should think that this intuition has such an exceptional epistemic status. But in the present context, the more important point is that even if (as moral irrealists would maintain), this intuitive norm lacks a rational ground, that would not be enough to suggest that we should abandon the norm. Indeed, presumably Singer himself, if convinced that there is no rational ground for any of our ethical commitments, would not abandon all of those commitments. (Note that in the lay population, people who claim to reject moral objectivism still retain commitment to familiar moral principles and treat them much the same as people who embrace moral objectivism [Nichols 2004].)

None of this is to say that the presence of a rational ground is epistemically irrelevant. If there is a rational ground for some ethical norm, then *that's* the best reason for sustaining the norm. But if the irrealist is right that none of our ethical norms has any ultimate rational ground, that doesn't mean that we are obliged to abandon our norms.

[19] Norms that are *not* inferentially basic are undercut if they depend on factual or inferential errors. To take an obvious example, many people who oppose abortion do so on the basis of factual beliefs (e.g. about whether the fetus has a soul). If the facts that underpin their normative conviction are known to be false, this destroys the warrant for retaining the norm.

suffering children" provides a clear example. That norm is widespread in our culture. And it is plausibly a result of our emotional reactions to suffering and not from a process of rational inference from more basic principles.[20]

Ethical conservatism maintains that the fact that our ethical norms are not ultimately justified does not eradicate their normative legitimacy. This accords well with the classical sentimentalist tradition, which draws a deep analogy between aesthetics and ethics. In both cases, our normative judgments are a function of human nature – it's because of the emotional nature we happen to have that we make the aesthetic and ethical judgments that we do. In addition, classical sentimentalism maintains that the fact that emotions are at the bottom of our ethical and aesthetic lives doesn't undermine the legitimacy of either ethical or aesthetic norms (see Gill 2007).

In sum, ethical conservatism accords a special status for normative commitments that are rooted in human emotion and are not inferentially dependent on other norms or facts. These norms can retain their normative legitimacy even though they might result from patently arational processes. The primary reason for accepting ethical conservatism is that it provides a natural way to preserve commonsense morality in the face of its emotional, arational origins.[21] And the analogy with aesthetics provides a kind of existence argument for how emotion-based norms can retain their normative legitimacy .

6. Competing considerations

The retributive norm is entrenched – it is widespread, inferentially basic (section 2) and rooted in a basic human emotion (section 3). Since the retributive norms is entrenched, the ethical conservatism sketched above entails that the retributive norm is not undermined by the absence of rational justification. The bare retributive norm thus retains an initial normative legitimacy, even in the face of its arational, emotional basis. But of course the standing of entrenched norms is not unassailable. The view is *conservative* not *reactionary*. Being an ethical conservative does not entail rigid deference to all entrenched norms. There can be competing moral considerations that lead us to think it right to reject an entrenched norm.

Some entrenched commitments do not face obvious competing moral considerations: e.g. "help suffering strangers" seems broadly consistent with the remainder of our moral concerns.

---

[20] We have a large body of entrenched norms, including norms against theft, incest, and harming innocents, as well as norms promoting retribution, reparations, and special obligations to family. Some normative ethicists might presume that one of our entrenched norms can be the foundation for all of ethics. These are large issues in philosophical ethics, and I can't do justice to them here. But I see no reason to expect that one of our entrenched norms will serve to fund all of the rest of our cherished ethical commitments. Nor do I see any overwhelming normative reason to impose such a monistic assumption on our ethical theorizing. As a result, I propose to proceed with an ethical conservatism that grants an initial legitimacy to each entrenched norm.

[21] One might maintain that the category of entrenched norms is too broad, and that some classes of entrenched norms should not be granted special status of initial normative legitimacy. For example, one might maintain that entrenched norms based on disgust don't merit this status. That is an interesting view, but it would require an argument to motivate it. In the absence of such an argument, I'll proceed with the broad category that grants the special status to all entrenched norms.

Other entrenched norms do face competing moral considerations. Consider first an example quite apart from punishment – organ donation. There continues to be significant resistance to organ donation, with cultural variations in the extent of resistance (see, e.g., Braun & Nichols 1997).There *is* something repellant about the practice of having one's child disemboweled after death. It's likely that there is (or was) an entrenched norm against defiling the body of one's deceased child by extracting her organs. The majority of Westerners now accept the propriety of organ donation.[22] And we've done so because of a competing ethical consideration – an enormous benefit to another person. In light of the fact that a deceased child's organs can save another person's life, we have overcome our entrenched norm against allowing someone to root around in our dead children's bodies.

Are there competing moral factors when we look to our retributive norm? Of course. *Ceteris paribus* it's wrong to harm others. Retributivism involves *intentionally harming* others. This is the core of the familiar complaint against retributivism – retributivism promotes intentionally harming someone without getting any (other) benefit out of it (Duff 2008). This is obviously an important challenge to retributivism. To begin to meet it, I first want to suggest that in Western philosophical ethics, we have come to fetishize harm, as if it trumps any other ethical considerations. But there are plausibly cases in which it can be appropriate to cause harm to someone in the service of other non-harm-based concerns. Say that I promised my grandmother that I would visit her grave every Memorial day, but that when the time comes, I need to wake my sleeping wife to get the car keys. Waking someone from sleep sets back their welfare, probably more so than a punch on the arm. But we regard waking someone under those circumstances as permissible. Or, to take a pervasive case, most readers will agree that it is okay to tax the successful bachelor farmer even if the government services he gets are not commensurate with the amount of taxes he pays. Many programs that our tax dollars support do not provide any benefit for the successful bachelor farmer: public broadcasting, NEH, NEA, humanitarian aid. The bachelor farmer might well not want to relinquish a penny for any of these programs. At best, they are of no use to him; at worst, he might find them despicable. In any case, taking his money for these services is an unequivocal harm to him. But that doesn't make us think it wrong to tax him. For, at least in some cases, like humanitarian aid, we think that moral considerations outweigh the harm to him.

Of course, the fact that it can be acceptable to cause a harm in the service of some competing ethical norms doesn't show that this is okay in the case of retributive punishment. It might be that harming really does globally override the retributive norm. But the nature of the issue is obscured by focusing on the severity of the punishment. Thus Duff speaks of the worry that our practice of punishment "seeks to inflict significant hardship or pain." (2008). The issue is further obscured by the fact that discussions of retributivism typically revolve around punishing criminals in a large-scale society. But the bare-retributive norm doesn't say anything directly about technique or state sponsored punishment. It just says that wrongdoers should be punished.

I want to argue that there are cases of punishment in which the fact that we are producing a harm isn't, by itself, enough to override the legitimacy of the retributive norm. The fact that we're harming someone isn't enough to undermine the judgment that the wrongdoer should be punished. To soften the ground, though, I want to return to the case of organ donation.

In "The Wisdom of Repugnance", Leon Kass maintains that our natural revulsion at

[22] The Gallup Organization, Inc. 2005 National Public Opinion Survey on Organ Donation

various medical procedures is ethically revelatory: "Repugnance… revolts against the excesses of human willfulness, warning us not to transgress what is unspeakably profound." (Kass 1997, 20). This repugnance suggests to us, according to Kass, that there is something wrong with, *inter alia*, in vitro fertilization, cloning, and organ donation. Kass has been rightly criticized for his faith in repugnance as an ethical indicator (e.g., Harris 1998). While "wisdom" no doubt inflates the epistemic value of repugnance, there is something to the idea that repugnance is not to be ignored. As noted earlier, there is a natural resistance to the idea that it's okay to have one's child disemboweled after death. In our culture, we have come to accept the propriety of organ donation. But note that the benefits of organ donation are tremendous. Imagine that the benefits were smaller, say, that by organ transplantation, the net gain would be a reduction in the number of sniffles. In that case, it would scarcely be worth suspending the norm against disemboweling dead children. I would not sanction harvesting the organs of my dead child to help relieve someone's runny nose. Kass might be wrong to view repugnance as wise. But it's not like our felt resistance to organ donation counts for *nothing*.

Recall where we are – the claim is that we have an entrenched commitment to the retributive norm and this endows the norm with a legitimacy that isn't eradicated by the fact that the norm has an arational source. Does the fact that retribution involves intentional harm count as sufficient reason to overturn the retributive norm? The point of the organ donation example is that the stakes can make a big difference to how we think about competing ethical considerations. To evaluate the competing considerations in the case of punishment, it will be important to avoid vexed issues about criminality and incarceration. Instead, let's return to our economic games. In these games, people think that those who behave unfairly should have money deducted from their fund. For instance, many people say that one should pay $.25 to have $.75 deducted from the unfair player. And virtually everyone says that if it doesn't cost the punisher any money, he should deduct money from the unfair player. Here we have a sharp case of competing factors. The retributive norm says *punish* (by $.75), and this competes with the fact that we would be setting back B by $.75 (with no benefit). In this competition, does the loss of $.75 to B suffice to overturn the norm that wrongdoers should be punished? Does our commitment to retributivism count for so little? 75 cents? It is an affront to commonsense ethics that our commitment to retribution is so cheaply bought off. Of course, as the harms increase in scale, the competition can become more challenging. At some point, no doubt, the retributive norm does get overridden in virtue of the competing considerations. But if I'm right that $.75 doesn't buy off the retributive norm, then there is no general rejection of retributivism forthcoming from the fact that it involves intentionally harming someone.

Competing considerations about intentional harm might tip the balance against harsh punishments, but that doesn't justify *globally* overturning our entrenched commitment to retribution. When we consider institutionalized forms of punishment, like incarceration or the death penalty, it might be appropriate to oppose those techniques on a variety of grounds. These practices might be excessively severe; the practices might unfairly skew to certain racial groups; the practices might demand a higher standard of evidence of guilt than is typically obtained. There are numerous reasons one might oppose capital punishment or incarceration. But, if the line of argument here is right, the fact that the punishment is retributive is not itself a reason to oppose it.[23]

[23] In my defense of brute retributivism, I have avoided appealing to any benefits of retributivism. For to appeal to benefits of retributivism leaves one open to that charge that retributivism is

7. Conclusion

Despite its ubiquity in ordinary life, the retributive norm has resisted rational justification. This leads many to reject the legitimacy of retributivism. But unless we take on controversial metaethical assumptions like moral realism, we are bound to accept some basic moral norms without justification. For the vast bulk of ordinary moral thought likely emanates from arational, emotional sources. Rather than conclude that this invalidates ordinary moral thought, we can reject the assumption that, in order to carry normative legitimacy, a norm must be justified (or capable of justification). Some norms retain normative legitimacy even if they have no independent justification. I've suggested that the norms that have this special status are those that are widespread, rooted in emotions, and inferentially basic. The bare retributive norm falls in this class. As a result, the fact that we can't justify the norm doesn't defeat the norm. This leaves open a number of issues about punishment - how to punish, what to punish, who should punish. Indeed, it remains an open question whether, all things considered, we should sustain retributivism. But we shouldn't take the lack of justification as a reason to abandon our brute retributivism.

References:

Baier, K. 1955. Is Punishment Retributive? *Analysis*, 16, 25–32.

Bedau, H. 1978. Retribution and the Theory of Punishment, *Journal of Philosophy*, 75, 601-20.

Benn, S. 1967. Punishment. In P. Edwards (ed.), *The Encyclopedia of Philosophy*, vol. 7, 29–36. New York: Macmillan and the Free Press.

Berlin, A. & Brettler, M. 2004. *The Jewish Study Bible.* Oxford: Oxford University Press.

Blackburn, S. 1998. *Ruling Passions.* Oxford: Oxford University Press.

Blair, J. 1995. A Cognitive-Developmental Approach to Psychopathy. *Cognition* 57,1-29.

Bolton, G. & Zwick, R. 1995. Anonymity versus punishment in ultimatum Bargaining *Games and Economic Behavior* 10, pp. 95-121

Boyer, P. 2000. Evolution of the Modern Mind and the Origins of Culture. In P. Carruthers and A. Chamberlain (eds.), *Evolution and the Human Mind*. Cambridge, UK: Cambridge University Press, 93-113.

Braithwaite, J. 1999. Restorative Justice: Assessing Optimistic and Pessimistic Accounts. In M Tonry (ed.), *Crime and Justice: A Review of Research*, vol. 23. Chicago: University of Chicago Press, 241-367.

Braun, K., & Nichols, R. 1997. Death and dying in four Asian American cultures: A descriptive study. *Death Studies*, 21, 327-359.

---

being justified by *nonretributive, consequentialist* considerations (Mackie 1982, 208; Bedau 1978, 616). However, when we turn to the broader question about whether to sustain the retributive norm in light of other ethical considerations, we must consider the costs and benefits of the retributive norm. And the benefits appear to be quite significant. In particular, the retributive norm likely plays a crucial role in facilitating cooperation (see Gaus 2011).

Breiter H., Gollub R., Weisskoff R., Kennedy D., Makris N., Berke J., Goodman J., Kantor H., Gastfriend D., Riorden J., Mathew R., Rosen B., Hyman S. 1997. Acute Effects of Cocaine on Human Brain Activity and Emotion. *Neuron 19*, 591-611.

Buhrmester, M., Kwang, T., and Gosling, S. 2011. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?' *Psychological Science* 6, 3-5.

Carlsmith, K., Darley, J. & Robinson, P. 2002. Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology*, 83, 284–299.

Carlsmith, K. 2008. On Justifying Punishment: The Discrepancy Between Words and Actions. *Social Justice Research.*

Clarke, S. 1728. *A Discourse on Natural Religion,* 7th edition.

Cottingham 1979. Varieties of Retributivism. *The Philosophical Quarterly, 29.*

Darwin, C. 1872. *The expression of the emotions in man and animals.* London: John Murray.

Dawes, C., Fowler, J., Johnson, T., McElreath, R. & Smirnov, O. 2007. Egalitarian Motives in Humans. *Nature* 446, 794-796.

Duff, A. 2008. Legal Punishment. *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), E. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/legal-punishment/>.

Fehr, E. & Fischbacher, U. 2004. Third party punishment and social norms. *Evolution and human behavior* 25, 63-87.

Fehr, E. and Gächter, S. 2000 Cooperation and punishment in public goods experiments. *American Economic Review* 90, 980–994.

Fehr, E. and Gächter, S. 2002. Altruistic punishment in humans. *Nature* 415, 137–140

Fiala, B., Arico, A., and Nichols, S. forthcoming. "On the Psychological Origins of Dualism: Dual-process Cognition and the Explanatory Gap." In E. Slingerland & M. Collard (eds.) *Creating Consilience: Issues and Case Studies in the Integration of the Sciences and Humanities.* Oxford University Press.

Frank, R. 1988. *Passions Within Reason.* New York: W. H. Norton.

Freud, S. 1927/1961. *The Future of an Illusion.* Translated by J. Strachey. New York: Norton & Co.

Gaus, J. 2011. Retributive Justice and Social Cooperation. In M. White (ed.), *Retributivism: Essays on Theory and Practice*. Oxford: Oxford University Press, forthcoming.

Gibbard, A. 1992. *Wise Choices, Apt Feelings.* Cambridge, MA: Harvard University Press.

Gill, M. 2007. Moral Rationalism vs. Moral Sentimentalism: Is Morality more like Math or Beauty? *Philosophy Compass* 2, 16–30.

Gill, M. and Nichols, S. 2008. Sentimentalist Pluralism, *Philosophical Perspectives* 18, 143-163.

Greene, J. 2008. The Secret Joke of Kant's Soul. In W. Sinnott-Armstrong (ed.), *Moral Psychology*, Vol. 3. Cambridge, MA: MIT Press.

Haidt, J., Bjorklund, F., and Murphy, S. 2000. Moral Dumbfounding: When Intuition Finds No Reason. Unpublished manuscript. University of Virginia.

Hampton, J. 1984. The Moral Education Theory of Punishment. *Philosophy and Public Affairs* 13, 208-38.

Harman, G. and Thomson, J. 1996. *Moral Relativism and Moral Objectivity*. Cambridge, MA: Blackwell.

Harding, A. 1966. *A Social History of English Law*. Baltimore, MD: Penguin Books.

Harris, J. 1998. *Clones, Genes, and Immortality*. Oxford: Oxford University Press.

Hopfensitz, A. and Reuben, E. 2009. The Importance of Emotions for the Effectiveness of

Social Punishment. *Economic Journal* 119, 1534-1559.
Hume, D. 1739/1964. *A Treatise of Human Nature.* Oxford: Clarendon Press.
Husak, D. 1992. Why Punish the Deserving? *Nous* 26, 447-64.
Joyce, R. 2002. *The Myth of Morality*. Cambridge: Cambridge University Press.
Kass, L. 1997. The Wisdom of Repugnance. *New Republic*, June 2, 17–26
Lemerise, E. and Dodge, K. 2008. The Development of Anger and Hostile Reactions. In Michael Lewis et al. (eds.), *Handbook of Emotions*, 3rd Edition. Guilford Press, 730-741.
Mackie, J. 1977. *Ethics: Inventing Right and Wrong*. New York: Penguin.
Mackie, J. 1982. Morality and the Retributive Emotions. *Criminal Justice Research* 1, 3-10.
Maine, H. 1861. *Ancient Law.* London: John Murray.
Menninger, K. 1968. *The Crime of Punishment*. New York: Viking Press.
Mikula G. 1986. The experience of injustice: toward a better understanding of its phenomenology. In H. Bierhoff, R. Cohen, & J. Greenberg (eds.), *Justice in Social Relations*. New York: Plenum, 103–24.
Miller, W. 1990. *Bloodtaking and Peacemaking.* Chicago: University of Chicago Press.
Moore, M. 1987. The Moral Worth of Retribution. In F. Schoeman (ed.), *Responsibility, Character, and the Emotions*. Cambridge, UK: Cambridge University Press.
Murphy, J. 1985. Retributivism, Moral Education and the Liberal State. *Criminal Justice Ethics* 4, 3-11.
Nadelhoffer, T., Heshmati, S., Kaplan, D. & Nichols, S. (forthcoming). Folk Retributivism: In Theory and in Practice.
Nichols, S. 2002. On the Genealogy of Norms. *Philosophy of Science* 69, 234–55.
Nichols, S. 2004. *Sentimental Rules.* New York: Oxford University Press.
Nichols, S. forthcoming. Ethics and Debunking.
Nichols, S., Timmons, M., and Lopez, T. forthcoming. Ethical Conservatism and the Psychology of Moral Luck. In M. Christen et al. (eds.) *Empirically Informed Ethics.* Springer.
Pillutla, M. & Murnighan, J. 1996. Unfairness, anger and spite. *Organizational Behavior and Human Decision Processes*, 68, 208-224.
Prinz, J. 2007. *The Emotional Construction of Morals.* Oxford, UK: Oxford University Press.
De Quervain, D., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., and Fehr, E. 2004. The Neural Basis of Altruistic Punishment. *Science* 305, no. 5688, 1254-1258.
Rawls, J. 1955. Two Concepts of Rules. *The Philosophical Review* 64, 3-32.
Sanfey, A., Rilling, J., Aronson J., Nystrom L., Cohen, J. 2003. The neural basis of economic decision making in the Ultimatum Game. *Science*, 300, 1755-1758.
Scholl, B. 2007. Object persistence in philosophy and psychology. Mind & Language, 22, 563-591.
Shafer-Landau, R. 1996. The Failure of Retributivism. *Philosophical Studies* 82, 289-316.
Singer, P. 2005. Ethics and Intuitions. *Journal of Ethics* 9, 331-352.
Sinnott-Armstrong, W. 2006. *Moral Skepticisms.* Oxford: Oxford University Press.
Smith, A. 1790/1982. *The Theory of Moral Sentiments.* Indianapolis, IN: Liberty Classics, 1982.
Sommers, T. forthcoming. *Relative Justice*. Princeton University Press.
Sperber, D. 1996. *Explaining Culture*. Cambridge, MA: Blackwell.
Stein, E., John Pankiewicz, Harold H. Harsch, Jung-Ki Cho, Scott A. Fuller, Raymond G. Hoffmann, Marjorie Hawkins, Stephen M. Rao, Peter A. Bandettini, and Alan S. Bloom. 1998. Nicotine-Induced Limbic Cortical Activation in the Human Brain: A Functional

MRI Study. *American journal of psychiatry* 1998 155, 1009-1015
Strawson, P., 1962. Freedom and Resentment. *Proceedings of the British Academy* 48, 1-25.
Street, S., 2006. A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies* 127, 109-66.
Timmons, M. 1999. *Morality Without Foundations: A Defence of Ethical Contextualism*. New York: Oxford University Press.