

A “Slice of Cheese” – A Deterrence-based Argument for the International Criminal Court

Jakob von Holderstein Holtermann

Abstract: Over the last decade, theorists have persistently criticised the assumption that the International Criminal Court (ICC) can produce a noteworthy deterrent effect. Consequently consensus has emerged that we should probably look for different ways to justify the ICC or else abandon the prestigious project entirely. In this paper I argue that these claims are ill founded and rest primarily on misunderstandings as to the idea of deterrence through punishment. They tend to overstate both the epistemic certainty as to and the size of the deterrent effect necessary in order to thus justify punishment. I argue that we should in general expect reasonably humane punitive institutions to lead to better consequences than if we abolish punishment entirely. And I show that, contrary to widespread assumption among critics of the ICC, we should not expect the conditions characteristically surrounding mass atrocity to undermine this presumption. Properly understood, the ICC equals adding another “slice of cheese” to our comprehensive crime preventive system modelled along the lines of James Reason’s Swiss cheese model of accident causation and risk management. Undoubtedly, some future perpetrators will elope through the holes in this layer too, but others will be deterred.

Keywords: The International Criminal Court, mass atrocity, punishment, deterrence, prevention, the “Swiss cheese model”

E debbasi considerare come non è cosa più difficile a trattare, né più dubbia a riuscire, né più pericolosa a maneggiare, che farsi capo ad introdurre nuovi ordini. Perché lo introduttore ha per nimici tutti quelli che delli ordini vecchi fanno bene, et ha tepidi defensori tutti quelli che delli ordini nuovi farebbono bene.¹

– N. Machiavelli, *Il Principe*

Oui, mais il faut parier : cela n'est pas volontaire, vous êtes embarqué. Lequel prendrez-vous donc ? Voyons. Puisqu'il faut choisir, voyons ce qui vous intéresse le moins.²

– B. Pascal, *Pensées*

Introduction

On December 31 2000, after then US president Bill Clinton had put his (later withdrawn) signature on the Rome Statute on the International Criminal Court, he issued the following statement:

¹ “And it ought to be remembered that there is nothing more difficult to take in hand, more perilous to conduct, or more uncertain in its success, than to take the lead in the introduction of a new order of things. Because the innovator has for enemies all those who have done well under the old conditions, and lukewarm defenders in those who may do well under the new.”

² “Yes; but you must wager. It is not optional. You are embarked. Which will you choose then? Let us see. Since you must choose, let us see which interests you least.”

“I believe that a properly constituted and structured International Criminal Court would make a profound contribution in deterring egregious human rights abuses worldwide.”
(Clinton 2000)

Through these words, Clinton joined in with human rights organisations and politicians alike who celebrated the prospects of finally bringing an end to impunity for mass atrocities through the creation of the ICC.

Among punishment theorists in the Academy, however, this hopeful enthusiasm has had severe difficulties in taking root. To the extent the ICC has in fact found (hesitant) defenders it has almost without exception been in the retributivist quarters. Any attempts to justify the ICC consequentially on preventive grounds have met with widespread scepticism. In particular, the claim that the ICC can be justified on grounds of deterrence has met with massive counterarguments; arguments that it seems are generally considered knock-down.

I believe, however, that most of these arguments are unconvincing and rest primarily on misunderstandings as to the idea of deterrence through punishment. In particular, they tend to overstate both the epistemic certainty of consequences and the size of the deterrent effect necessary in order to thus justify punishment. No doubt, despite the creation of the ICC we will remain saddeningly far from fulfilling the urgent post-Holocaust imperative of “Never again”. But I do believe that Clinton was basically right: a properly constituted and structured International Criminal Court will in due time be able to deter *enough perpetrators to justify its costs*. All things considered, I believe we are currently more justified in making this claim than we are in making the opposite. Thus, to consequentialists, knowing what we know now, establishing the ICC seems to be the right thing to do. In this article I try to show why.

In focusing thus on the deterrent effect of punishment, I do not commit myself to any claim that punishment only serves preventive purposes. In particular, I do not exclude the possibility that punishing perpetrators of mass atrocities through a suitably constituted international criminal court will in addition provide a more respectful way of treating victims’ calls for justice than do other attempts to deal with an atrocious past including the currently fashionable model of truth and reconciliation commissions.³

Focusing on deterrence I also leave out several aspects that properly belong in a comprehensive discussion of the possible crime preventive effects of establishing the ICC. In particular, I do not discuss (in any great detail) the possible incapacitative or rehabilitative effects of punishment through the ICC. Nor do I discuss what has been described as the “displacement function of law”, i.e. the ability of punishment to “remove some of the temptation to retaliate” (Gardner 1998, p. 31).

As to incapacitation and rehabilitation, this omission is excused by the fact that these aspects play little or no role in current discussions of punishment of mass atrocities. And apparently they do so for a good reason. If we have in fact managed to put perpetrators of mass atrocities on trial it is usually because they are definitively no longer in a position to offend on that scale. In this sense, mass atrocities are typically “one-off crimes”. Omitting the displacement function can, on the other hand, only be excused on account of limited space. I believe that this aspect of punishment deserves far greater attention than it has received so far – both in punishment theory generally and in the context of the ICC and mass atrocities in particular.

Even within discussions of deterrence proper important aspects are left out here. In particular, I do not enter the classical ethical dispute over the failure of deterrence theory to guarantee the rights of the innocent. In spite of its classical status in traditional punishment theory, this line of criticism plays a limited role in actual discussions of the ICC. I shall only briefly indicate in the conclusion why this might be so.

Despite this heavy trim, there remains, as we shall see, ample material for discussion.

³ For a thoughtful and convincing argument to that effect, see Brudholm (2008).

I proceed as follows: in part I, I discuss a general epistemological objection against attempts to justify punishment on grounds of deterrence. In part II, I consider three specific arguments that have been launched in order to show that the so-called “domestic analogy” breaks down, i.e. that even if punishment can be justified in domestic contexts any such justification cannot be transferred from the domestic context of “ordinary” crime to the international context of mass atrocities. In the conclusion I sum up the results of my findings and indicate where discussion could go from here.

Part I: Deterrent justifications of punishment as such⁴

The underdetermination objection

The first criticism aims generally at the possibility of justifying punishment *as such*, that is, regardless of it being used in domestic or international settings, on grounds of deterrence. This argument is rarely addressed openly in discussions of the ICC but it nevertheless lurks in the background of the writings of several critics (cf. e.g. Drumbl 2007, p. 171; Mégret 2001, p. 202; Tallgren 2002). This leaves the impression that any talk of deterrence is fundamentally flawed and discussion in the particular context of mass atrocities and the ICC proceeds only for sake of argument and on account of charity. I shall therefore discuss the general objection at some length before proceeding to the particular question of punishment of mass atrocity.

This argument centres on the alleged lack of empirical evidence in support of the claim that punishment can work thus as a deterrent. For example a recent survey of punishment theories over the last half-century cites the absence of empirical evidence as the main reason for the decline of consequentialist theories of punishment since the 1970s:

“By the 1970s, it was clear that the social sciences could not then, or in the foreseeable future, give empirical preventive theories much empirical support. The social sciences could not, that is, say what effect, if any, statutory penalties, rehabilitation, exemplary punishment, or even incapacitation would have on the crime rate (much less whether those effects would repay the cost).” (Davis 2009, p. 84)

This claim can reasonably be conceded – if only for sake of argument and pending further specification. The factors influencing criminal behaviour arguably do seem abundant and the number of possible combinations thereof indefinite. In addition we have the apparently insurmountable practical and ethical difficulties of testing hypotheses on these matters satisfactorily.

And it would seem that any such difficulties of justifying punishment are multiplied many times over in the international context of mass atrocities where even more unknown factors are added to the cost-benefit calculus. Thus, as one ICC critic puts it:

“It is not easy to estimate how likely the preventive effect of the international system is. There are no grounds to exclude the possibility of such an effect. Neither is there evidence in its favour.” (Tallgren 2002, p. 569)

The central question is, however, what inferences are justified in response to these facts. Davis, for instance, concludes that we should look for different ways to justify punishment altogether (2009, p. 85) and argues that such justifications should rely on *conceptual* instead of empirical relations. Tallgren on the other hand appears to find punishment of international crime unjustified on any grounds:⁵

⁴ The argument propounded in this first part of the article is developed more fully in (Holtermann 2009a).

⁵ She suggests instead a rather conspiratorial genealogical explanation as to why the ICC has nevertheless come into existence:

“Prevention is cited simply because of the void of alternatives, the rational ones. ... International criminal justice comes close to a religious exercise of hope and perhaps of deception.” (Tallgren 2002, p. 561)

As will become clear, however, there are good reasons to resist these discouraging conclusions.

To punish or not to punish, tertium non datur – a confusion of epistemological standards

The character of the objection is basically epistemological. The critics assume the sceptical stance specifically with regard to beliefs as to the possibility of deterring potential offenders through the use of punishment. The available evidence, they claim, fundamentally underdetermines any beliefs on this issue. As such, the objection renders topical traditional epistemological discussions that are not always dealt with satisfactorily in punishment theory.

First, we should notice carefully what is or rather what is not warranted by this observation. In particular, we are not justified in positively claiming that punishment is *not*, on balance, a cost beneficial way of dealing with crime. That is, we are not justified in inferring from these epistemological difficulties that the costs of punishing criminals will, on balance, lead to *worse* consequences than if we abolish punishment entirely.⁶ In this sense, the epistemological position with regard to beliefs on crime prevention is entirely analogue to the sceptical position generally:

”Strictly, skepticism is a matter of doubt rather than of denial. The skeptic is strictly not one who denies the validity of certain claims, but one who questions, if only initially and for methodological reasons, the adequacy of our grounds for holding them.” (Strawson 1987, p. 2)

In epistemological parlour, then, what is justified is only to withhold judgement on the matter in question. The epistemological difficulties warrant only that we refrain from forming beliefs on the matter of crime prevention through punishment.

However, as indicated by Strawson, thus withholding our judgement is an artificial epistemological construct. Human beings are not only belief-forming creatures. We continuously act and, notably, interact with the surrounding world. I can, if I try really hard using Cartesian methodological doubt, withhold judgement as to whether breathing or withholding my breath will be more beneficial for my general well-being. But I cannot *not* do either. I will inevitably either breathe or withhold my breath, and my choice of path will, equally inevitably, have some highly tangible consequences for my well-being.

In this sense, withholding judgement is an abstraction; an epistemological luxury we cannot afford in everyday life. Being physical creatures our beliefs will inevitably convert into action at some point. This illustrates a more general point already acknowledged by Descartes on the interim character of sceptical doubts:

“Perhaps its task is to naturalize, to exclude from the political battle, certain phenomena which are in fact pre-conditions for the maintenance of the existing governance; by the North, by the wealthy states, by wealthy individuals, by strong states, by strong individuals, by men, especially white men, and so forth.” (Tallgren 2002, pp. 594-595)

⁶ This point was explicitly conceded by Tallgren in the first of the above quotes. Admittedly, criticism of preventive theories of punishment has occasionally been launched under the motto “nothing works” (cf. Lipton, Martinson, & Wilks 1975). Interpreted in this sense, the criticism does form a positive proposition (though formulated in the negative) as to the likely consequences of punishment. As indicated by the title, however, this conclusion is primarily aimed at the perhaps overly ambitious rehabilitation programs of the 1960’s and 70’s. Interpreted as a general claim that punishment does not, all else equal, tend to reduce crime, however, most commentators agree – as I will get back to below – that the evidence does not warrant this conclusion.

“In the meantime, I know that no danger or error will result from my plan, and that I cannot possibly go too far in my distrustful attitude. *This is because the task now in hand does not involve action but merely the acquisition of knowledge.*” (Descartes 1996, p. 15, my emphasis)⁷

But as Descartes implicitly recognises he cannot remain thus inactive forever. Soon the six days of meditation will be over, and short of indubitable knowledge in a world of events relentlessly unfolding he will then have to act on those beliefs that he finds, on balance, are the most justified.

And the same point holds good in punishment theory – perhaps even more so than in theoretical philosophy since, here, the task in hand patently does involve action and not (only) the acquisition of knowledge. Regardless of the epistemological difficulties in ascertaining the exact effects of punishment on crime, we will inevitably have to either establish punitive institutions or refrain from so doing.

This reveals a basic premise of moral life on consequentialist conditions: While omniscience as to the consequences of our actions would of course be preferable, epistemological difficulties in this regard do not amount to a decisive argument against consequentialist theories. In and of themselves they only serve to remind us of the painful human condition of being fallible creatures equipped with a conscience.

In the absence of knowledge beyond some absolute threshold of doubt, then, the central issue becomes one of relative or comparative certainty between mutually exclusive beliefs. And in the case of punishment theory this means asking whether, on balance, we are *more* justified in believing that punishing crime will have better consequences than we are in believing the opposite. Which of the two makes up the more reasonable default presumption? In other words, we have to determine which view holds *the burden of proof*.

There is absolutely nothing odious in thus lowering the standard of doubt and proceeding on less than complete or “scientific” certainty if we have no alternative. We do it constantly in countless areas of life, e.g. perhaps most notoriously in civil law, where the burden of proof is met if a proposition has been proven to be more probable than its negation. Surely we do not for that reason consider such judgements unjustified, nor do we attempt to escape the epistemological problems by taking Davis’ route and pretend that the judgements *really* express conceptual truths rather than somewhat doubtful empirical matters (Davis 2009, p. 85).

It could perhaps be objected that the example of civil law is badly chosen precisely because *criminal* law (in which context the question of the justifiability of punishment belongs) traditionally works with a wholly different and much stronger burden of proof, i.e., that of proving guilt *beyond reasonable doubt*. However, the contradiction dissolves when we take a closer look. When subjected to epistemological limitations, it is, other things equal, reasonable to act in accordance with a cautionary principle, i.e., to err on the side of caution. This principle presupposes, however, that one such side can be identified. In the individual criminal trial most agree that this is easily done. The nightmare of a single wrongful conviction by far outweighs the presumably infinitesimal loss in crime preventive effects and also even the grievances of victims and relatives of a single wrongful acquittal. Hence, the famous Blackstone’s formulation: “[I]t is better that ten guilty persons escape than that one innocent suffer.” (Blackstone 1860, pp. Book 4, *358) In the more general discussion pro and con having a punitive institution at all, things are not that simple. Other things equal, it is not immediately clear whether it is better to live in a society where a lot of crime is committed by criminals that could have been deterred by a lawful threat of punishment, than it is

⁷ Ancient sceptics make an analogue point. Thus, e.g., Sextus Empiricus rejects the claim that his epistemological views should lead to a life of inaction. Our epistemological limitations only influence the modus of our so-called knowledge claims. Thus, Pyrrhonian scepticism, he claims, is perfectly consistent with having beliefs about appearances as long as we do not dogmatise: “Adhering, then, to appearances we live in accordance with the normal rules of life, undogmatically, seeing that we cannot remain wholly inactive.” (Empiricus 1933, book I, p. 23)

to live in a society where a lot of fundamentally unjustified punishment is inflicted.⁸ Thus, it seems that when it comes to the fundamental question of justifying the punitive institution as such, we have no alternative to the weaker standard of proof used also in civil law. Which of course is not to say that we cannot, and indeed should not, hold on to the traditional stronger standard of reasonable doubt when determining criminal guilt in court.

Punishment does prevent crime

Anthony Ellis has made an analogue point in a recent article (2009), and he argues, in my view convincingly, that opponents of punishment must bear the burden of proof against proponents. The main reason is simply that human beings are by and large instrumentally rational actors. That is, their behaviour can be understood and explained all predominantly as attempts to achieve goals that they perceive to be desirable or, conversely, to avoid states of affairs that they consider undesirable.⁹ Absent this hypothesis we would generally be banned from using *causa finalis* in the explanation of human behaviour, in which case nothing we now know as psychology would be possible. It is so to speak the possibility condition of psychology.¹⁰

Granting this presumption of instrumental rationality, and adding that punishment induces suffering that human beings, other things being equal, strive to avoid¹¹ it seems prima facie reasonable that people will generally be deterred, at least to some degree, from engaging in behaviour that is criminalised if they are (aware that they are) threatened with punishment for so doing, and if this threat is (perceived to be) credible, i.e. if there is some likelihood of detection.

This would indeed help explain conspicuous phenomena like the dramatic increase in crime level in Denmark during World War II following immediately upon the detention of the entire Danish police force by the German occupying forces (cf. Trolle 1945). To the polemically minded any such observations are, of course, vulnerable to *post hoc, ergo propter hoc*-accusations, but it does seem to be a reasonable explanation that the sudden disappearance of a credible threat of punishment had something to do with it.

Several criminological studies have further corroborated the general claim that punishment deters. To mention just one prominent meta-study, criminologist Daniel S. Nagin concludes thus: “[T]he accumulated evidence on deterrence leads me to conclude that the criminal justice system exerts a substantial deterrent effect.” (2000, p. 359)¹² In other words, in the absence of indubitable knowledge the *more* reasonable claim surely seems to be that punishment exerts a deterrent effect.

Exaggerated doubts and “supposed universal idiocy”

⁸ This could perhaps be challenged on a version of the principle of double effect according to which the unjustified punishments would be intended and hence worse than the excess crimes that would only be foreseen. This, however, is a retributivist rationale, and therefore inapplicable here as the criticism presently under consideration claims that the preventivist line of argument fails even on its own terms.

⁹ Which of course is not to say that people always know what is good for them, nor that they cannot have mutually exclusive desires. It is not even to say that people cannot want something to happen that they consider bad in itself. But when they do so, it is usually because they consider it *instrumentally* good, i.e. expedient.

¹⁰ In Ellis’ words: “Without such a presumption, it would be impossible to understand human affairs at all.” (2009)

¹¹ Of course there are exceptions to this general rule. Dostoyevsky’s portrait of the painter Mikolai in *Crime and Punishment* provides one famous example of an inversion of traditional human sensibilities. More generally, such inversion is found e.g. in the phenomenon of voluntary penance. Gang members that consider e.g. prison convictions marks of honour provide yet another example of such inversion.

¹² Cf. also e.g. Walker (1991, pp. 13-20) and Wilson (1983, ch. 7).

Apparently this much is conceded even by deterrence-critics like Davis (2009, p. 85).¹³ He denies, however, the second step necessary in order to justify punishment generally on grounds of deterrence; i.e. he denies that we should have any knowledge as to whether the benefits of thus reducing crime actually *outweigh* the costs of establishing them:

“Neither Ellis nor any other empirical preventivist has been able to answer that question [whether the extent of the deterrence justifies the costs involved]. None knows whether punishment does deter enough to be justified by its deterrent effect.” (Davis 2009, p. 86)

Again, this is surely true if by “know” we mean true belief established beyond all scientific doubt. Since the claim is basically counterfactual we would ideally have to compare total societies that use punishment with (like) societies that do not in order to establish it properly. However, all known human cultures use some kind of sanctions for rule breaking (Brown 1991, p. 138), and chances are rather limited that social scientists will be given an opportunity to conduct a realistic large-scale experiment in which a society does not do that any time soon. Hence, the comparison admittedly remains, in some sense, speculative.¹⁴

However, as already mentioned we are not here (necessarily) trying to establish scientific beliefs beyond some absolute standard. We are trying to determine which claim should bear the burden of proof in view of our epistemic limitations. And I must admit that I find it hard to believe that Davis and others honestly find themselves in serious doubts on this issue. Of course we cannot rule out entirely *the possibility* that abandoning punishment completely will lead to greater human happiness, than what is achieved by establishing punitive institutions roughly resembling those we would currently find it relevant to consider. However, the very absence of just one example throughout human history of a society successfully abandoning punishment makes it rather hard to take that possibility seriously. On the contrary, inference to the best explanation of this fact arguably does seem to be that such cultures simply are not very viable.¹⁵

The actual uncontroversiality of this assumption is further supported by the fact that it is rather hard to find serious abolitionists in punishment theory (and not just proponents of drastic reductions in the use of punishment). Even advocates of so-called restorative justice¹⁶ who are often associated with abolitionism agree that restorative justice processes must take place, as one advocate puts it, “in the shadow of the axe” (Braithwaite 2002, p. 36). Anything else, they admit, would be unrealistically utopian.

Perhaps Davis and likeminded critics simply exaggerate their doubts methodologically in order to establish a critical point. If so I find the approach misguided. It is generally hard to see why

¹³ Though Davis finds this deterrent effect to be “conceptual” rather than empirical. It is doubtful if this distinction makes any sense. For an extended discussion, see Holtermann (2009a).

¹⁴ Although game theorists have done extensive work explicating the role of sanctions and the consequences of their absence in relations to the achievement of socially desirable ends. Here mathematical modelling (cf. e.g. Axelrod 2006) as well as experimental research (cf. e.g. Camerer 2003) seems to suggest that the presence of informal or formal sanctions is necessary for the kind of cooperative institutions fundamental to social organization to be viable. Davis appears, however, to dismiss any such evidence, and I shall for sake of argument not build my case on it. (I am grateful to Pelle Guldborg Hansen for bringing this to my attention.)

¹⁵ This universal agreement indicates that perhaps political philosophers since Hobbes have not been entirely wrong in their depiction of the state of nature as an intolerable and ultimately untenable situation. One reasonable explanation for the universal use of sanctions for wrongdoing could easily be that the absence of punishment implies the absence of a local monopoly of violence that in turn implies the breakdown of social institutions and norms and finally a descent into a state of nature.

To illustrate, the earlier mentioned detention of the Danish police force during WWII was immediately countered by the creation of municipal corps of watchmen that in spite of limited means and powers managed to take the edge of the exploding crime (Christensen 2001).

¹⁶ Restorative justice ordinarily refers to a theory of criminal justice that emphasises repairing the harm caused by crime rather than administering “hard treatment” on the offender, and that couples this general aim with a heavy presumption in favour of reaching it through informal deliberative stakeholder processes (i.e. victim-offender mediation, conferencing, circles). For a discussion, see e.g. Holtermann (2009c).

theories that justify punishment on empirical grounds should be denied access to such highly general but perfectly reasonable beliefs simply for want of them being “scientific” – whatever that means. Insisting that consequentialist theories should be put to such a hard test brings to mind a line of criticism that John Stuart Mill emphatically warned against:

“It is truly a whimsical supposition that, if mankind were agreed in considering utility to be the test of morality, they would remain without any agreement as to what is useful, and would take no measures for having their notions on the subject taught to the young, and enforced by law and opinion. There is no difficulty in proving any ethical standard whatever to work ill, if we suppose universal idiocy to be conjoined with it...” (Mill 1987 [1871], p. 296)¹⁷

Short of thus supposing “universal stupidity”, then, I take it that there are no serious reasons for doubting that the beneficial effects of punishment in any society we find it relevant to consider outweigh the costs of establishing it.

A different angle – Buridan’s Ass revisited

The epistemological objection can, however, be given a slightly different interpretation which deserves attention before dealing with the specific arguments launched against punishing international crimes. Again, Davis delivers the input when he specifies the reason why he believes that the social sciences by the 1970’s could not give preventive empirical theories much support:

“If even relatively crude tuning of penalties to empirical consequences is in practice impossible, empirical preventive theories cannot justify punishment as an institution, *much less choosing any institution of punishment over others, or choosing one punishment over another*, except in some counterfactual world in which we would know much more than we do know here. A theory of punishment should be more practical than that.” (Davis 2009, p. 84, my emphasis)

Basically, this portrays preventive theory as falling prey to the practical fallacy of Buridan’s Ass which dies from starvation between two equally attractive heaps of hay because it cannot find a decisive reason for choosing one over the other. On this interpretation, then, *even if* we have conceded that, on balance, we are better off punishing, then we would still wind up not punishing if we justify punishment solely on empirical grounds – simply because we would be incapable of choosing between several possible institutions of punishment that look equally attractive from our epistemologically limited perspective.

Two things can be said in reply: first, this objection arguably repeats the overestimation of the epistemological difficulties at play. Thus, we are not *completely* ignorant as to the punitive institutions at our disposal. History and our general knowledge of the sensibilities of the human psyche manifestly root out some of the more extreme possibilities. For example, sanctioning serious wrongdoing by some kind of social exclusion presents itself as a more natural choice than, e.g., do tickling.¹⁸ And social science arguably narrows the options even further. What the evidence ultimately underdetermines then is only what Nagin describes as largely incremental policy questions: “comparatively small changes grafted unto the status quo system of enforcement and penalties.” (2000, p. 366) Properly described then the dilemma is actually reasonably manageable:

¹⁷ Charles S. Peirce makes a similar albeit more general criticism against the entire idea of building philosophy on a Cartesian concept of methodological doubt: “Let us not pretend to doubt in philosophy what we do not doubt in our hearts.” (1868, p. 140)

¹⁸ In fact, Brown recounts that in all known human cultures the use of sanctions “include[s] removal of offenders from the social unit – whether by expulsion, incarceration, ostracism, or execution” (1991, p. 138)

“While it is my view that the evidence points to the entire enterprise having a substantial impact, predicting the timing, duration, and magnitude of the impact of incremental adjustments in enforcement and penalties remains largely beyond our reach.” (Nagin 2000, p. 366)

Second, critics would do well to remember that Buridan’s Ass is indeed a satirical figure. It takes a fairly (pardon the expression) dumb ass to be paralysed by the prospect of two similar looking heaps of hay. Analogously, even though, as Davis notes, a theory of punishment should indeed be practical, it is surely acceptable for it to leave some level of arbitrariness in choosing one particular punitive system over others.

On the whole, punishment theory is, in this respect, in no different position from numerous other areas of human life where we try to balance costs and benefits in large-scale complex systems in less than perfect epistemic conditions. Thus (to mention one prosaic but illustrative example), in the case of industrial production of chemical substances nobody knows *precisely* how best to balance the risks to human health and environment with considerations of e.g. innovative capability, industry competitiveness and human progress. Even after ruling out some obviously implausible extremes our limited knowledge of the countless factors involved and their incalculable future ramifications surely leaves us with a large number of equally attractive specific solutions as to the balancing of these costs. However, neither in this area (nor in countless areas like it) do we habitually suffer the unfortunate fate of Buridan’s Ass.¹⁹ On the contrary, we choose one particular balance over all the many others available. For example, in 2006 the EU passed the comprehensive so-called REACH-regulation that deals with the Registration, Evaluation, Authorisation and Restriction of Chemical substances. REACH requires, for instance, that

“[t]he supplier [of a substance] shall provide the recipient at his request with a safety data sheet [...] where a preparation [...] contains in an individual concentration $\geq 1\%$ by weight for non-gaseous preparations and $\geq 0,2\%$ by volume for gaseous preparations at least one substance posing human health and environmental hazards...” (REACH 2006, p. 108)

Empirical considerations clearly did not dictate the choice of *this particular* regulation. But it would plainly be absurd to claim, for that reason, that the fundamental justification for the regulation cannot be empirical but must be *conceptual* to use Davis’ terminology. That would equal labelling *conceptual* the decision of any creature wise enough to escape the sorry fate of Buridan’s Ass.²⁰

We should have no reason to view the matter differently in the case of crime and punishment. Neither in this area does a certain level of arbitrariness on grounds of epistemic uncertainty leave us paralysed. Thus I wholly agree with Ellis:

“[N]o more can be required than that it [a political community] make a reasonable, good-faith estimate of the balance [where the costs of punishment are proportional to its benefits], and try to stay within it. [...] So long as a community does this, it will be justified in using punishment.” (Ellis 2009)

¹⁹ Or hardly ever. But when we do, it is usually not due to our epistemological limitations but, rather, to the lack of political courage or to the political process reaching a deadlock.

²⁰ Other examples abound of such highly detailed, but ultimately arbitrary regulations, which can nevertheless only be justified, fundamentally, on empirical grounds. Consider, for instance, the so-called Eurocode regulation: “a set of common standards containing the European calculation methods to assess the mechanical resistance of structures or parts thereof” (European Commission 2009). Surely, the only available justification of such regulation is empirical rather than conceptual.

(This, of course, is not to say that there is no morally relevant difference between dealing with crime and punishment and e.g. chemicals and building construction. It is only to say, that any difference there might be is not principally epistemological.)

Once the manifest exaggerations have been weeded out our admittedly less than perfect knowledge as to the exact consequences of punishing crime does not amount to an argument against justifying punishment on empirical grounds. If one is basically sympathetic to the empirical outlook established empirical knowledge justifies punishment roughly within the framework provided by modern rules of law – which of course still leaves ample room for such incremental adjustments which I, following Nagin, would label almost all current policy questions.

Part II: The international context – attacking the domestic analogy

The discussion so far has had no particular bearing on punishment of the crimes of genocide, war crimes and crimes against humanity *qua international* crimes. As mentioned, I have discussed the general epistemological objections to punishment as such because a general disbelief in the preventive effects of punishment lurks in the background in some of the critical writings on the ICC.

For the most part, however, critics of attempts to justify the ICC primarily on grounds of deterrence actually do seem to accept that burden of proof, which I have so far argued should rightly be placed on their shoulders. That is, they accept, if only for the sake of argument, at least the theoretical possibility of such a deterrent effect of punishment. Instead they challenge the so-called domestic analogy, that is, the claim that this deterrent effect can be transferred from the stable context of “ordinary” domestic crime to the unstable context of the extraordinary international crimes of genocide, war crimes and crimes against humanity. Rhetorically, then, deterrence critics have focused their efforts on breaking down the continuity between these two spheres, and they have done so by pointing to the existence of particular circumstances surrounding the occurrence of international crimes that renders it difficult or even impossible to establish the wanted deterrent effect in that particular context.

I shall focus on this line of criticism in the remaining part of the article.²¹ Before proceeding, however, I should perhaps recount what is at issue here. Thus, the claim that an institution like the ICC can be justified on deterrence grounds:

- is *not* a claim that everybody will be deterred all the time;
- is a claim that some will be deterred some of the time; and
- is a claim that it will deter a sufficient number of potential perpetrators to justify the costs of producing this effect.

Ordinary people in extraordinary and complex circumstances

Critics of the ICC tend to emphasise that these mass crimes are not committed solely, or even typically by a few evil monsters or abnormal psychopaths. On the contrary, genocide, war crimes and crimes against humanity generally presuppose mobilising the masses which is possible only through the establishment of a set of extraordinary socio-psychological circumstances. Here is Mark A. Drumbl:

”Violence becomes normalized when neighbors avert their gaze, draw the blinds, and excitedly move into a suddenly available apartment. This broad public participation, despite its catalytic role, is overlooked by criminal law, thereby perpetuating a myth and a deception. The myth is that a handful of people are responsible for endemic levels of violence. The de-

²¹ Several critics have presented criticism along roughly congenial lines. For ease of presentation, I focus primarily on Drumbl (2007), Tallgren (2002) and Wippman (1999) in the following. For an overview, see e.g. Mennecke (2007).

ception [...] involves hiding the myriad political, economic, historical, and colonial factors that create conditions precedent for violence.” (Drumbl 2007, p. 172)²²

Drumbl presents these considerations as an argument why the ICC cannot be justified on grounds of deterrence. Apparently, the idea is that because more than “a handful of people are responsible” punishing only this “handful” will not exert a deterrent effect. But as we shall see, this argument rests on a fundamental misunderstanding of the claimed deterrent effect of punishment.

On the factual side, Drumbl is undoubtedly right as to the large number of conditions precedent for mass violence. In this he is supported by the rich literature within both genocide studies generally and perpetrator studies in particular. Both these areas have indeed contributed to an unmasking of myths and provided a more nuanced and richer image of the contexts in which atrocities take place and of those who commit them. For instance, a classical study lists, among other things, the following socio-psychological factors that history have shown jointly to constitute what is described as a “continuum of destruction” (cf. Staub 1989, p. ch. 2):

- i) Difficult life conditions: e.g. economic problems (inflation, depression, etc.), widespread violence and social disorganisation which lead to an experience of threat to physical safety and psychological self;
- ii) Cultural and personal preconditions: e.g. ingroup-outgroup differentiation, devaluation of outgroup, orientation to authority, monolithic (vs. pluralistic) culture; and
- iii) Societal-political organization: e.g. authoritarian or totalitarian system, discriminating social institutions, institutions capable of carrying out mistreatment.

In addition, the less than sensational psychological profile of the majority of those who commit mass atrocities has been confirmed thoroughly in a wide variety of perpetrator studies from Milgram’s psychological experiment on obedience to authority (cf. 1974) to more recent historical case studies like Browning (1992) and Goldhagen (1996).

I see no reason to doubt these and like conclusions in general, but it is unclear what, if any, bearing they should have on deterrence theory. The complexity of the contexts in which mass violence takes place and the ordinary character of most perpetrators can perhaps be problematic if one subscribes to an expressionist or communicative theory of punishment like the one proposed e.g. by R.A. Duff (2001) according to which we punish in order to communicate community censure to the perpetrator. On this account it does perhaps present some difficulties to punish people who have been placed in a complex context in which they could hardly have acted differently.²³

But Drumbl explicitly presents these facts as an argument against attempts to defend punishment on grounds of deterrence, and I simply cannot see why this should be so. First, in and of themselves the facts only tell us that the veneer of civilisation is regrettably thin. We are all – or almost all – capable of committing the most horrifying deeds under the right (i.e. wrong) circumstances. As Glaucon tells Socrates in *The Republic* most if not all of us will regrettably seize the chance and perpetrate if the Ring of Gyges²⁴ is passed around widely:

“And this we may truly affirm to be a great proof that a man is just, not willingly or because he thinks that justice is any good to him individually, but of necessity, for

²² Tallgren mentions roughly analogue factors in her discussion of the possibilities of establishing general prevention through an international criminal justice system (2002, pp. 570-576). But in fact she explicitly admits that they have no bearing on the possible deterrent effect of punishment (p. 576).

²³ Duff obviously does not think so (cf. 2009).

²⁴ The Ring of Gyges grants the one who wears it the power to become invisible, and Glaucon uses it in order to discuss whether ordinary people would act in accordance with the precepts of morality if they did not have to fear the consequences of their actions.

wherever any one thinks that he can safely be unjust, there he is unjust." (Plato 1901, p. 38, my emphasis)

If, however, one justifies punishment on deterrent grounds, this does not imply any claim as to the extraordinarily evil or depraved nature of the perpetrator. On the contrary, it only presupposes that human beings are capable of doing evil in specific contexts, and that we can and should try to oppose this unfortunate predisposition by influencing the structure of incentives.

Second, it appears to me that Drumbl confuses necessary and sufficient conditions in the argument. In order to sort out this confusion I suggest that instead of talking unqualified of *causes* or *responsibility* for mass violence we introduce the concept of a NESS-test, which was developed in order to deal with the intricate issue of causality in law.

Following Wright (2001)²⁵, NESS is acronym for necessary element of a sufficient set, and the concept captures that what is usually referred to as *causes* or as being *responsible* for an effect, rarely brought about that effect single-handedly. If, for instance, I hit the light switch and the lights turn on we would usually say that I *caused* or was *responsible* for turning on the light. If, however, the light switch had, say, been disconnected from the electric system or there had been no light bulb in the lamp, my doing so would not have turned on the lights. Thus, being connected to the electric system, having a light bulb in the lamp and countless other conditions constitute the set which in conjunction with my hitting the light switch are sufficient for turning on the lights. They are all NESS-conditions.

This concept of a NESS-condition is particularly useful in the analysis of complex events like genocide, crimes against humanity and war crimes. Thus it arguably applies to each of "the myriad political, economic, historical, and colonial factors" which Drumbl mentions and which Staub and other scholars of genocide and perpetrator studies try to outline more systematically. Each of these factors, empirical studies seem to tell us, does indeed have to be present in order for mass violence to occur. But interestingly the NESS-test also applies to the absence of a credible threat of punishment. Genocide etc. simply do not take place unless the Ring of Gyges has been passed around widely. To my knowledge, throughout history genocides have never taken place unless a credible threat of punishment from local authorities has been absent at the time of perpetration.²⁶ Only in conjunction with this condition, it seems, is the set of factors sufficient.

If indeed this is so, it is hard to see why a legal system whose primary aim it is to prevent mass violence should be banned from establishing punitive institutions that aim primarily at removing this single factor known to be a NESS-condition for the occurrence of that sort of violence. Thus aiming at one factor simply does not commit the legal system to any claim that none of the other factors belong in the picture. If any myth or deception is perpetuated in this regard it is entirely due to the presence of retributivistic rationales in the courtroom.

What the story of the myriad factors *does* tell us, on the other hand, is that in crime prevention we do not have to think of punishment as our *only* chance of preventing mass crime. Instead, we should use a "broad-spectrum drug" in the sense that we should target each of those factors that genocide and perpetrator studies tell us pass the NESS-test, and we should do so with means tailored specifically to each of them.

It would perhaps be helpful to think of crime prevention generally in terms of the so-called Swiss cheese model, which James T. Reason (cf. e.g. 2000) has introduced in the field of risk analysis and risk management in complex human security systems like aviation control, nuclear plants, hospitals etc. In these areas, the main goal is also preventive, i.e., to prevent accidents, and when accidents occasionally do happen it is also best analysed in terms of NESS-conditions, i.e., a number of factors combine simultaneously in such a way as to constitute a set sufficient for the adverse effect. On these conditions, Reason invites us to envision security systems as successive layers of Swiss cheese, where each layer is designed to remove one particular kind of NESS-condi-

²⁵ Wright attributes the concept to Hart and Honore (1959).

²⁶ Or unless the threat of punishment has been annulled by a more immediate threat of force. The distinction is not important here.

tion, i.e., one of the factors known to be necessary for an accident to occur. The *Swiss cheese*-part enters the picture because, in complex human systems, it is impossible to design perfect layers. Unfortunately, none of the NESS-conditions can be removed permanently. The challenge, then, is to design the layers in such a way as to prevent that “the holes in many layers momentarily line up to permit a trajectory of accident opportunity” (Reason 2000, p. 769).

Applied to the field of mass violence each slice of cheese represents an (alas too imperfect) attempt to remove a NESS-condition on the “continuum to destruction” (e.g. severe economic problems, social disorganisation, ingroup-outgroup differentiation etc. - cf. above). Establishing the ICC simply adds another layer at the end of this continuum by attempting to end impunity by threatening punishment. Undoubtedly, this layer will also have plenty of holes in it (partly for reasons to be discussed in the following) but this is no reason not to add it to our preventive system. Thus, we get the following version of James Reason’s Swiss cheese model:

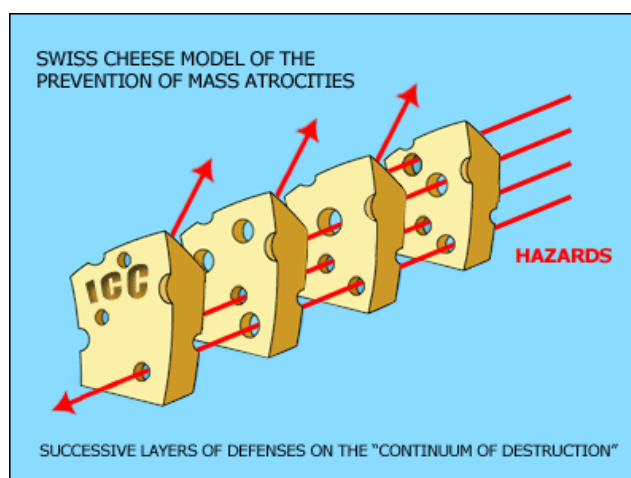


Fig. 1 The Swiss Cheese model of the prevention of mass atrocities (cf. also Reason 2000)

Instrumental rationality in times of mass violence?

A related though distinct criticism of the ICC concerns the necessary presumption of instrumental rationality among perpetrators if talk of deterrence is to make any sense. Several theorists have challenged the applicability of this presumption in the contexts of mass violence. For example, Drumbl remains unconvinced that punishment of international crimes will effectively deter:

“because deterrence’s assumption of a certain degree of perpetrator rationality, which is grounded in liberalism’s treatment of the ordinary common criminal, seems particularly ill fitting for those who perpetrate atrocity.” (Drumbl 2007, p. 171)²⁷

Much depends, of course, on how one reads this claim. But interpreted as a radical claim that instrumental rationality should generally be totally absent in times of mass violence, I simply cannot see why this should be so. As earlier noted by Ellis, “Without such a presumption [of instrumental rationality], it would be impossible to understand human affairs at all.” (2009)²⁸

²⁷ See also, Martha Minow: “Individuals who commit atrocity on the scale of genocide are unlikely to behave as “rational actors”, deterred by the risk of punishment.” (1998, p. 50) Parallel passages can be found also e.g. in Tallgren (2002, p. 584) and Mégret (2001, p. 203).

²⁸ Though independently developed and structured, the following argument is fundamentally congenial to Ellis’ convincing discussion in (2009).

On the contrary, the very possibility of conducting perpetrator studies testifies to the existence of such rationality – if rightly understood. It shows that, however perverted and abhorrent, perpetrators' entire motivational structure is indeed pervious to scientific enquiry. This is so because as human beings perpetrators continuously make some sense of their own behaviour. They justify their actions to themselves, and if this justification is not immediately obvious to everyone watching (e.g. in the shape of the barrel of a gun pushing the perpetrator forward), it can be meticulously reconstructed through empirical studies. This is exactly what Staub, Browning and others have shown.²⁹ They have shown the step-by-step procedure that renders it possible for perpetrators gradually to justify their increasingly perverted actions. Indeed, these studies have shown what makes these actions appear to the perpetrators to be the only *rational* thing to do. Thus, perpetrators of mass violence are no doubt rational actors in a sense sufficiently rich to be compatible with the fundamental assumptions behind deterrence theory.

The criticism of the rational actor-assumption can, however, be interpreted more plausibly as a claim that, though rational, perpetrators of mass violence are typically embedded in a motivational structure that renders it impossible for them to react on the threat of punishment the way rational actors would normally do. And this is so because such perpetrators usually live hand-in-mouth. Each day they struggle for survival in some of the most callous and dangerous environments on earth. In order to describe their life conditions we might safely ignore the metaphorical “as” of Matthew Arnold’s famous poem “Dover Beach”:

“And we are here as on a darkling plain
Swept with confused alarms of struggle and flight,
Where ignorant armies clash by night.”

With a gun in your back you need something roughly as convincing pointing in the opposite direction in order not to keep moving. And an ICC that has no police, is placed far away in The Hague, and which is probably going to take several years in order to gather sufficient evidence to present an arrest warrant, seems to be entirely incapable of presenting a threat of that calibre.

On the whole, the difficulty of establishing a credible threat of punishment is often presented as a main reason for scepticism as to the ICC. Here is Wippman:

”When the various motivations for attacks on civilians are combined – the desire to defend one’s community, hatred of the other side, a belief that civilians are essentially indistinguishable from combatants and therefore a threat, and directions from political and military leaders encouraging such attacks – it is not surprising that a slight risk of future prosecution will not have a major deterrent effect.” (1999, p. 479)³⁰

No doubt rendering the threat of punishment credible is one of the major challenges for the ICC. However, for two reasons this need not be a decisive problem. First, there is a risk of overstating the point. Even during the perpetration of mass atrocities it does occasionally happen that perpetrators find the time and surplus energy to take precautionary measures as to the possibility of future sanctions. As Wippman continues:

”Serb forces in Kosovo routinely wore black ski masks when engaged in ethnic cleansing; as the prospect of NATO control over Kosovo loomed larger, Serb forces intensified efforts to conceal mass graves and hide evidence of criminal conduct.” (1999, p. 480)³¹

²⁹ Which of course is not to say that such studies cannot disagree internally as to their specific theories of perpetrator motivation.

³⁰ Cf. also Minow: “[I]t is not irrational to ignore the improbable prospect of punishment given the track of record of international law thus far.” (1998, p. 50)

³¹ Ellis refers to analogue behaviour on behalf of operators of Nazi concentration camps as allied victory approached (2001, p. 111, n13)

Thus, it appears that, however remote, the surrounding world *does* stand a chance after all of influencing perpetrators behaviour – at least to some degree.

Second, there is (cf. the first Wippman-quote above) an outspoken tendency in these arguments to focus on foot soldiers, i.e., on rank-and-file acting under orders and/or in massively coercive social circumstances. And this is admittedly where the difficulties of reaching potential perpetrators with an external threat of punishment seem the most overwhelming. However, this focus is seriously misleading. The ICC is on all accounts a big-fish court. Its aim is not the individual perpetrator armed with a machete but those political and military leaders who conceived, initiated and ordered the mass atrocities. Thus, among probably tens of thousands of perpetrators actually involved in committing various mass atrocities in Congo, Uganda, the Central African Republic and Darfur, Sudan since 2002 the ICC-prosecutor has (as of May 29, 2009) initiated only thirteen (!) cases (ICC 2009). It is of course possible that once fully operable the ICC will indict more per conflict, but we should not expect the total number of inditees to exceed more than a tiny fraction of the total number of actual perpetrators of mass atrocities. Thus, to illustrate, the fully operable ICTY and ICTR indicted 161 and 90 persons respectively (Drumbl 2007, p. 170).³²³³

Political and military leaders are generally placed in circumstances where their contribution plays a decisive role as events unfold. Even though they surely cannot perform mass atrocities single-handedly their actions as leaders are undoubtedly necessary in order for them to take place. That is, they are NESS-conditions in the above sense. And there are several good reasons to suppose that, in contrast to foot soldiers, these leaders can, in the right circumstances, be deterred by the threat of punishment, i.e., that they are instrumentally rational in the sense desired.

First, unless of course we give in entirely to determinism and fatalism and deny ourselves any possibility of acting differently than what past events had in store (in which case we do not have to worry too much whether or not the ICC can be justified because it will inevitably either come to be or not according to its destiny), political and military leaders will typically act under less immediate pressure from the surroundings than do foot soldiers who are under orders and at gunpoint. As Payam Akhavan observes:

“Once mass violence has erupted, threats of punishment can do little to achieve immediate deterrence. However, the outbreak of such violence can be inhibited, and its resumption in postconflict situations prevented – because it often results from an elite’s deliberate political choices.” (Akhavan 2001, p. 10)

With the “confused alarms of struggle and flight” safely at a distance political and military leaders arguably seem to be left with some room for action in which the possibility of future punishment can play a role.

To this should be added an admittedly somewhat speculative and generalising psychological consideration, which nevertheless deserves to be taken seriously. It seems reasonable to suppose that in order to reach the top of the power hierarchy in those dictatorships and fascist states that are usually racked by mass atrocities (indeed to become a leader in any political community), it is generally necessary to possess an exceptionally high degree of instrumental rationality. It is hard to think of a better sign of mastery of unsentimental Machiavellian means-end reasoning than having reached a high-level position in politics. One generally does not get that far by living hand-in-mouth. It is done only by thinking strategically and by taking the long view. While political and military leaders certainly may not always have demonstrated a grasp of the categorical imperative it may safely be assumed that they master the hypothetical imperative considerably better than the majority.

³² For those entirely sceptical as to the possibility of the international community eventually bringing perpetrators of mass atrocities to justice, the numbers of the ICTY and ICTR should give some reason to pause. Of the 161 indicted by the ICTY only two remain at large as of May 2009 (ICTY 2009). Of the 90 indicted by the ICTR 13 remain at large (ICTR 2009).

³³ In fact, the ICTY initially even indicted some foot soldiers too. A strategy later to be abandoned on grounds analogue to those presented here (Akhavan 2001, p. 19).

It is therefore generally hard to see why the pitiful fate of e.g. Milosevic, Karadzic, and even long time fugitive Mladic should not leave an impression on at least some political and military leaders who are about to initiate mass atrocity and who find themselves in roughly analogue geo-political realities, e.g. being leaders of mid-size states, possibly even with powerful allies in the Security Council. For these reasons, I believe Akhavan is wholly right:

“Where leaders engage in some form of rational cost-benefit calculation, the threat of punishment can increase the costs of a policy that is criminal under international law. Leaders may be desperate, erratic, or even psychotic, but incitement to violence is usually aimed at the acquisition and sustained exercise of power.” (2001, p. 12)

If indeed the threat of punishment is credible we should therefore expect high-level leaders to take it into consideration before initiating mass atrocities.³⁴ And this contention is even further supported by an additional consideration to be presented in the following.

The peace over justice-argument

Several critics have, however, denied this conclusion. And they have done so on grounds that ultimately lead to yet another separate argument against the attempts to justify the ICC on grounds of deterrence. This line of argument takes as its point of departure the fact that historically warnings of future punishment for various kinds of mass violence have rarely proven themselves efficient:

”Actual experience with efforts at deterrence is not encouraging. ... Beginning in 1941, the United States and the United Kingdom issued a series of highly publicized warnings that violations of the laws of war would be punished and that superior orders would not be accepted as a defense. ... Similarly, in the former Yugoslavia, the Security Council and various individual states repeatedly warned combatants that those committing atrocities would eventually be prosecuted. But ... there is no empirical evidence of effective deterrence in either case.” (Wippman 1999, p. 474)

On the contrary, critics claim that the threat of punishment often *works counter* to its express intention because it tends to prolong armed conflict and perhaps even intensify mass killings rather than deter from them. This has for instance been claimed in criticisms of the ICC-decision to prosecute leaders of The Lord’s Resistance Army in Uganda. Similarly, the indictment of Sudan’s president has been accused by the African Union of jeopardising the difficult peace process in the region (BBC 2008), and it has also led the former U.S. Special Envoy for Sudan to warn of “a disaster in the making” (Natsios 2008).

Thus understood the critics’ claim is not simply that the ICC is an impotent yet ultimately harmless endeavour but rather that it is in fact a dangerous institution that should be avoided at all costs.

Before replying to this general line of criticism it gives rise to one important comment. It should be strongly emphasised that this entire argument comes at a price, which may be too high for many deterrence-critics to pay. In effect, it decisively dismantles any charge along the rational actor-criticism discussed above as to the inevitable impotence of the ICC. It is plainly incoherent to hold, on the one hand, that because of the inherent complexities of mass atrocities any threat of punishment issued by the ICC will be incapable of reaching perpetrators in the “confused alarms of struggle and flight”, *while simultaneously* claiming, on the other, that upholding this threat can actually prolong and even intensify these same crimes. Deterrence critics will therefore have to

³⁴ Even Saddam Hussein was deterrable. Thus, as testified by Hans Blix (2004), leader of the UN arms inspectors in Iraq, Saddam did indeed cooperate once he faced an immediate and credible threat. Unfortunately for him, it appears the war had already been decided on other grounds.

make a choice between the peace over justice-argument and the arguments discussed earlier in this second part of the article.

And in fact, I believe that the most sceptical voices as to the inevitable impotence of an international criminal court have already been proven wrong by the above (and other like) events since the ICC came into existence (which of course is not to say that these same events have established its omnipotence either). Properly understood, then, the peace over justice-argument implicitly confirms the soundness of the domestic analogy on the very point where the rational actor-line of criticism attempts to break it down.

This does not change the fact, however, that deterrence theorists continue to face the challenge (as, indeed, do proponents of punishment of mass atrocities on *any* grounds) that while indeed the threat of punishment can have an effect on unfolding events, there is a risk of this effect being highly adverse, viz., in the shape of a prolonging and intensifying of mass atrocities.

This line of criticism appears, however, to misunderstand fundamentally the rationale behind any claim as to the deterrent effect of punishment. That the Nazis, the parties to the Balkan wars, the LRA, etc. have all continued and perhaps even intensified their crimes despite a direct threat of punishment should generally come as no surprise to anybody discussing punishment on grounds of deterrence. First of all, granting the distinctive character of mass atrocities, warnings of the kind Wippman refers to simply came too late. As earlier noted, once erupted there is enormous inertia in mass atrocities. It is therefore, as Akhavan notes with specific reference to the former Yugoslavia:

“... unrealistic to suppose that the ICTY could have instantaneously deterred crimes in the midst of a particularly cruel interethnic war in the former Yugoslavia. Hastily erected bulwarks cannot be expected to save lives when the deluge has already begun. The threat of punishment – let alone an empty threat – has a limited impact on human behavior in a culture already intoxicated with hatred and violence.” (Akhavan 2001, p. 10)

To this should be added a more general point. If punishment has any deterrent effect at all, it is *ex hypothesis* predominantly because potential perpetrators *who, as of yet, have no outstanding business with the law*, refrain from perpetration from fear of punishment. People who have already trespassed but remain at large, on the other hand, have had their incentive structure irreversibly perverted. To them, the prospect of prosecution and punishment no longer constitutes a disincentive to perpetrate *because they are already due punishment*. On the contrary, such prospect obviously gives these perpetrators a strong incentive to resist arrest, if necessary through continued perpetration – hence the apt word *desperados*. On the whole, this is no different from national legal systems. An armed robber with the police close on her heels is more likely to do harm than one who has been granted safe conduct (see also Ellis 2009; Holtermann 2009b).

This reminds us that while the benefits of punishing crime (again, *ex hypothesis*) are measured primarily in potential crimes that ultimately remain uncommitted because of the threat of punishment, any costs incur almost exclusively on “the carrying out side” of the posing of this threat. From a preventivist point of view, anything we do in response to a particular crime *after* the perpetrator first trespassed is highly likely to be counterproductive with regard to that particular crime.³⁵ Once people have perpetrated they are in a sense written off as lost causes, i.e. as one of those on whom the threat of punishment *did not* work. We nevertheless carry out the threat despite its probable negative effects in each particular case in the belief that we can convince *others* of the credibility of the threat.³⁶ Punishments actually carried out, then, are assurances essentially directed to other people. This, in a nutshell, is the deterrent justification of punishment.

³⁵ Or almost anything. As mentioned in the introduction punishment is likely to serve a displacement function too.

³⁶ And this, of course, is what is generally unacceptable about the thought of deterrence to philosophers of a Kantian bent.

This also reveals that deterrence theories of punishment depend crucially, and in ways unparalleled by most other theories of punishment, on the permanency of the court issuing the threat. Punishments of individual perpetrators are capable of rendering the threat of punishment credible to other potential perpetrators only if: i) they are perceived by the potential perpetrators to be instantiations of rule-governed behaviour on behalf of the courts, and ii) such perpetrators suspect that their own intended crimes will fall under that same rule. In other words, deterrence presupposes that punishments in relation to a particular conflict can be regarded by potential perpetrators in conflicts yet to break out as *promises* of similar punishment if they should choose to transgress.

However, by definition it simply does not make sense for an ad hoc-tribunal to issue such a promise. Qua *ad hoc* it does not have the institutional mandate. Any rules it follows apply solely to the conflict mentioned in its founding statute. In this sense ad hoc tribunals act in settings that are analogue to Kant's imagined society that is to be resolved completely on the following day but which still has one murderer in custody (Kant 1996, p. 106). As we know, Kant's theory of punishment can survive in that setting. A deterrent theory, however, cannot.

Absent a permanent court it is of course possible and perhaps even likely that some future cases of mass atrocities would still have been prosecuted through the establishment of new ad hoc-tribunals. But in that case, the absence of a permanent court would nevertheless have been a sign that the international community simply did not want to be so obligated in the future. That is, it would not have wanted to *promise* punishment to future perpetrators. This is only done when any pre-existing regularities of sanctioning of mass violence are framed into a permanent legal system (with codified laws, courts, rules of procedure, etc.). Creating a standing international criminal court, then, is a particularly forceful way of enforcing and coordinating people's expectations as to this sanctioning. It is a speech act, which provides the institutional means in order to make this regularity common knowledge among potential perpetrators. Thus, we should expect the perceived likelihood of perpetrators being caught and prosecuted to have increased because of the creation of the ICC. Denying this would amount, in effect, to the claim that all potential perpetrators consider the signing of the Rome Statute and the creation of the ICC to be nothing but an empty promise; that they believe, for instance, that those who have so far been caught and prosecuted by the ICC would have been so in any event. Excluding facile statements of fatalism, it is hard to see which considerations would support such a categorical statement.

In fact, all of the above is deterrence theory 101, and, hence, the factors mentioned in the peace over justice-argument do not constitute a breakdown of the domestic analogy. On the contrary, they very much confirm it. The only difference is in scale and dimension. Potential dangers of upholding a threat of punishment against a dictator engaged in mass violence are far greater than those of upholding a threat against an armed robber. But so, obviously, are the potential gains in the shape of potential genocidaires backing down from initiating mass violence under the threat of punishment.

These considerations also explain why we should reject as unduly myopic claims that the international community should back down from its threat of punishment at each prospect of it affecting a prolonging of an ongoing conflict. It has been suggested that in such cases the international community should stand ready to either grant blanket amnesties on grounds of equity (May 2005) or to grant conditional amnesties in accordance with the model of the South African Truth and Reconciliation Commission (Roche 2005). But this is problematic first, because we have no guarantees that the granting of such amnesties will in fact halt ongoing conflicts. Thus, it has been persuasively argued that continued impunity will at least in some cases make it more likely that new atrocities will be committed (cf. Méndez 2001, p. 31). Second, to the degree amnesties will in fact help further peace in particular conflicts they plainly do so at the expense of undermining the credibility of the general threat of punishment issued by the court because the conditions for withdrawing this threat will be easily replicated by future perpetrators.³⁷ Thus, in effect, it would render the international community vulnerable to blackmail. And if I am right as to the existence of a deterrent effect of punishment, we should expect the temporary local benefits of giving in to

³⁷ For an extended discussion on this point, see Holtermann (2009b).

such blackmail to be outweighed by the lasting global benefits of having established a credible threat of punishment.³⁸

This, of course, is not to say that the ICC should never back down. If we have reason to believe that prosecution and punishment will lead to disastrous consequences, e.g. in the shape of nuclear war, it would surely be wrong to insist that the ICC-prosecutor should continue unaffected. Kantianism aside, ethical imperatives are undoubtedly best understood in an *other things being equal*-fashion. It is not only a sign of the arrogance of power but equally of moral perversion to subscribe to the motto of Roman emperor Ferdinand I: “Fiat iustitia, et pereat mundus” (Let there be justice, though the world perish).

It is only to say that an ICC that justifies punishment on grounds of deterrence must look beyond narrow national or regional interests when deciding on which cases to prosecute and ultimately punish. Otherwise it would defeat its own purpose. Thus, the sensible maxim for the ICC to follow vis-à-vis the peace over justice-argument is: “Fiat justitia si mundus non periet” (Let there be justice unless it will make the world perish).

Conclusion

Admittedly, the above discussion of the peace over justice-argument does not constitute a decisive argument against the concerns presented in that argument. Strictly, I have only illustrated how the potential adverse effects cited in it are wholly in accordance with the basic predictions of deterrence theory and thus that in themselves these effects do not constitute a counterargument to the theory. I have not proven, however, that the additional assumption necessary in order to justify punishment on deterrent grounds will also be satisfied, i.e. that punishing these perpetrators will also in fact have a deterrent effect on other potential perpetrators – let alone an effect sufficiently large to justify the costs of producing it.

And it could perhaps be objected that any manifest effects of the ICC occur only *in medias res* or even *ex post* the commitment of mass atrocities, i.e., when one has been so unfortunate as to be singled out in “the indictment lottery”. The potential perpetrators who have yet to actually trespass, on the other hand, will not take the ICC into account since the risk of being thus singled out is too small to take seriously.

While some potential perpetrators may undoubtedly think this way (and hence escape our preventive system through one of the holes in the Swiss cheese) others seem, however, to take the possibility of punishment more seriously. How else to interpret, for instance, the reluctance of the USA, China, Russia etc. to sign the Rome Statute, than that even super powers of their stature fear having members of their ruling elites indicted by the ICC? Granting the political realities of the international system they may not fear actual *punishment*. But it seems reasonable to suppose that

³⁸ Pertinent to determining whether or not the ICC can ultimately be justified on grounds of deterrence is of course also the issue of the financial cost involved, and in particular whether resources spent on the ICC could otherwise have been allocated to the other “slices of cheese” in our comprehensive atrocity preventive system, e.g. to debt reduction or economic aid programmes.

This is a difficult question involving notoriously tricky counterfactual reasoning, and dealing with it in great detail is beyond the scope of this work. But I will add a few comments in order to keep the worry in perspective. The 2009 budget appropriations for the ICC alone amounted to €101,229,900 (International Criminal Court 2008). In absolute figures, this is no doubt a large amount but it pales in comparison with the amount spent globally on economic aid programmes etc. each year (for instance, Denmark alone spent 20 times that amount in official development assistance in 2008 (OECD 2009)). In other words, even if we imagine, counterfactually, that *all* the money currently spent on the ICC would otherwise have been invested in various ways in sustaining the other “slices of cheese” (which by no means should be taken for granted), then we may safely assume any additional preventive effect resulting from this allocation to be only infinitesimal. Spending the money on the ICC, on the other hand, introduces a whole new “slice of cheese” to the many already in existence. A slice which, for the reasons presented in this article, we should expect to produce a more significant deterrent effect. (I am grateful to an anonymous reviewer for bringing this to my attention.)

they fear having the international embarrassment and political nuisance of an ICC indictment added to the existing system of political pressures. That is to say, they refuse to sign *because they estimate that signing will increase the likelihood of being indicted in the future, and they do not want the prospects of such indictment to enter their cost-benefit analysis in future decision situations*. And this is precisely what deterrence theorists, on the other hand, would like it to do.³⁹

However plausible, these considerations admittedly do not exclude definitively the possibility that any such effect will in fact be absent. And flatly denying this possibility would plainly be pulling the wool over our eyes. There can generally be no guarantees as to the consequences of our actions in complex, large-scale human systems, and thus neither in that of international crime and punishment.

This is indeed disturbing but as I have argued in the first part of this paper any such lack of certainty does not *per se* constitute an argument against deterrence theory. On consequentialist conditions, the mere existence of possibilities of adverse effects is not very interesting. In and of themselves they only serve to remind us of the painful condition of being both fallible and conscientious. Since choosing is imperative, i.e. since we cannot *not* either punish perpetrators of mass atrocities or refrain from so doing, the possibility in question must be rendered plausible by argument. We must be given positive reasons to take it seriously. And I believe that I have shown in this article that the arguments ordinarily presented to that effect fail.

Thus, I have argued, in the first part of the paper, that on balance we are more justified in believing that establishing reasonably humane punitive institutions will, *other things equal*, lead to better consequences than if we abolish punishment entirely. This assumption simply provides the best fit with our basic knowledge of human history and psychology. And I have shown in the second part of the paper that, contrary to widespread assumption among critics of the ICC, there is no reason to believe that the so-called domestic analogy breaks down in the context of international crime, i.e. that the particular conditions surrounding the perpetration of mass atrocities undermine the other things equal-proviso. In particular, I have shown:

- that we have good reason to presume that the absence of a credible threat of punishment serves as a NESS-condition for the outbreak of mass violence in the complex contexts in which it typically takes place. Thus, it is reasonable to suppose that the introduction of such a threat equals adding another “slice of cheese” in the Swiss cheese model of our combined efforts at preventing mass atrocities;
- that the rational actor assumption crucial to deterrence theory holds good also in contexts of mass atrocities, and that in spite of possible difficulties at reaching low ranking perpetrators we should expect members of the ruling elites to be capable of responding to credible threats of punishment; and finally
- that one of the criticisms most often heard, i.e. the peace over justice-argument, i) is inconsistent with the other two objections; and ii) if sound only confirms that the basic suppositions of deterrence theory holds good, even in the context of mass atrocity.

Hence, I conclude that until different arguments have been presented we are justified in assenting to Clinton’s words: “that a properly constituted and structured International Criminal Court can make a profound contribution in deterring egregious human rights abuses worldwide”.

This conclusion should be of interest even to those who do not believe that new arguments challenging this empirical prediction are forthcoming but who nevertheless remain sceptical as to whether this fact actually *justifies* the ICC. Any such scepticism would probably have to be presented along the more classical philosophical lines that as a punitive institution the ICC remain unjustified on deterrent grounds (as on consequentialist grounds generally) because it implies us-

³⁹ Whether or not, then, ultimately to label such deterrent effect a deterrent effect of *punishment* seems merely a matter of words. In any case, it would plainly be impossible for the ICC to be the cause of such nuisance did it not at least occasionally manage to actually carry out punishment. Hence, punishment is ultimately a necessary condition for any such deterrent effect resulting from the ICC. See also, Ellis (2009).

ing convicted perpetrators merely as a means to deter other perpetrators as a goal. Hence it fails to guarantee the rights of the innocent.

It appears, however, that this line of criticism is somewhat less convincing in the context of mass atrocities. By their very nature, mass atrocities are typically difficult to hide (cf. also Ellis 2009). In addition, they are usually subject to massive public interest and media coverage. Thus, it seems less than likely that an ICC-prosecutor should ever feel tempted to knowingly prosecute an innocent person for well-documented mass atrocities with hundreds or thousands of victims.

Die-hard retributivists will of course reply that such an ICC nevertheless remains unjustified because it still fails to guarantee the rights of the innocent *in principle*. We can always imagine possible worlds in which an ICC-prosecutor will in fact deceive the world opinion and knowingly punish an innocent dictator in order to secure some perceived good. I believe however that most people dedicated to finding ways of dealing with the disturbing phenomenon of mass atrocities will find this line of reasoning somewhat too theoretical to take very seriously.

Acknowledgments: I would like to thank Martin Vinding, Mikkel Munch-Fals, Thomas Søbirk Petersen, Pelle Guldborg Hansen, Kira Vrist Rønn, Inge Schiermacher, Stig Andur Pedersen, and in particular Frej Klem Thomsen, Jesper Ryberg and two anonymous reviewers at *Human Rights Review* for discussion and valuable comments on earlier versions of this paper.

References

- Akhavan, Payam (2001). Beyond Impunity: Can International Criminal Justice Prevent Future Atrocities? *The American Journal of International Law*, 95(1), 7-31.
- Axelrod, Robert M. (2006). *The evolution of cooperation* (Rev. ed.). New York: Basic Books.
- BBC (2008). AU rejects Bashir Darfur charges. Retrieved May 29, 2009, from <http://news.bbc.co.uk/2/hi/africa/7517393.stm>
- Blackstone, William (1860). *Commentaries on the laws of England : in four books*. Philadelphia: Childs & Peterson.
- Blix, Hans (2004). *Disarming Iraq* (1st ed.). New York: Pantheon Books.
- Braithwaite, John (2002). *Restorative justice & responsive regulation*. Oxford; New York: Oxford University Press.
- Brown, Donald E. (1991). *Human universals*. New York ; London: McGraw-Hill.
- Browning, Christopher R. (1992). *Ordinary men : Reserve Police Battalion 101 and the final solution in Poland*. New York, NY: Aaron Asher Books.
- Brudholm, Thomas (2008). *Resentment's virtue : Jean Améry and the refusal to forgive*. Philadelphia, Pa.: Temple University Press.
- Camerer, Colin (2003). *Behavioral game theory : experiments in strategic interaction*. N.Y. Princeton, N.J.: Princeton University Press.
- Christensen, Claus Bundgård (2001). På vagt i en lovløs tid. *Fortid og nutid*, 2(June), 91-109.
- Clinton, Bill (2000). Statement by the President, December 31, 2000. Retrieved May 29, 2009, from http://clinton4.nara.gov/library/hot_releases/December_31_2000.html
- Davis, Michael (2009). Punishment Theory's Golden Half Century: A Survey of Developments from (about) 1957 to 2007. *Journal of Ethics*, 13(1), 73-100.
- Descartes, Rene (1996). *Meditations on first philosophy : with selections from the Objections and replies* (J. Cottingham, Trans. Rev. ed / and a new introduction for this edition by John Cottingham. ed.). Cambridge: Cambridge University Press.
- Drumbl, Mark A. (2007). *Atrocity, punishment, and international law*. Cambridge: Cambridge University Press.
- Duff, Antony (2001). *Punishment, communication, and community*. Oxford ; New York: Oxford University Press.

- Duff, Antony (2009). Can We Punish the Perpetrators of Atrocities? In T. Brudholm & T. Cushman (Eds.), *The Religious in Responses to Mass Atrocity*. New York: Cambridge University Press, pp. 79-104.
- Ellis, Anthony (2001). What Should We Do With War Criminals? In A. Jokic (Ed.), *War crimes and collective wrongdoing: a reader*. Malden, MA; Oxford: Blackwell Publishers, pp. 97-112.
- Ellis, Anthony (2009). War Crimes, Punishment and the Burden of Proof. *Res Publica, forthcoming*.
- Empiricus, Sextus (1933). *Sextus Empiricus*. London, New York,: W. Heinemann; Putnam.
- European Commission (2009, January 22, 2009). Introduction to eurocodes. Retrieved May 29, 2009, from http://ec.europa.eu/enterprise/construction/internal/essreq/eurocodes/eurointro_en.htm
- Gardner, John (1998). Crime: in Proportion and in Perspective. In A. Ashworth, A. Von Hirsch & M. Wasik (Eds.), *Fundamentals of Sentencing Theory: Essays in Honour of Andrew Von Hirsch*. New York: Oxford University Press, pp. 31-52.
- Goldhagen, Daniel Jonah (1996). *Hitler's willing executioners : ordinary Germans and the Holocaust*. New York: Alfred A. Knopf.
- Hart, H. L. A., & Honore, Anthony Maurice (1959). *Causation in the Law*: pp. xxxii. 454. Clarendon Press: Oxford.
- Holtermann, Jakob v. H. (2009a), Caring About How the World Happens to Be - Reply to Davis, Unpublished manuscript.
- Holtermann, Jakob v. H. (2009b). The End of 'the End of Impunity'? The International Criminal Court and the Challenge from Truth Commissions. *Res Publica, forthcoming*.
- Holtermann, Jakob v. H. (2009c). Outlining the Shadow of the Axe - On Restorative Justice and the Use of Trial and Punishment. *Criminal Law and Philosophy*, 3(2 / June), 187-207.
- ICC (2009). Situations and Cases; All Cases. Retrieved May 29, 2009, from <http://www.icc-cpi.int/Menus/ICC/Situations+and+Cases/Cases/>
- ICTR (2009). Status of Cases. Retrieved May 29, 2009, from <http://69.94.11.53/default.htm>, "Cases", "Status of Cases"
- ICTY (2009). Fugitives. Retrieved May 29, 2009, from <http://www.icty.org/sid/10010>
- International Criminal Court (2008). Resolution ICC-ASP/7/Res.4: Programme budget for 2009. Retrieved May 29, 2009, from http://www.icc-cpi.int/iccdocs/asp_docs/Resolutions/ICC-ASP-ASP7-Res-04-ENG.pdf
- Kant, Immanuel (1996). *The metaphysics of morals*. Cambridge: Cambridge University Press.
- Lipton, Douglas, Martinson, Robert, & Wilks, Judith (1975). *Effectiveness of correctional treatment : a survey of treatment evaluation studies*. Springfield: Praeger.
- May, Larry (2005). *Crimes against humanity : a normative account*. Cambridge, UK ; New York: Cambridge University Press.
- Mégret, Frédéric (2001). Three Dangers for the International Criminal Court: A Critical Look at a Consensual Project. *Finnish Yearbook of International Law*, XII, 193-247.
- Méndez, Juan E. (2001). National Reconciliation, Transnational Justice, and the International Criminal Court. *Ethics and International Affairs*, 15(1), 25-44.
- Mennecke, Martin (2007). Punishing Genocidaires: A Deterrent Effect or Not? *Human Rights Review*, 8(4), 319-339.
- Milgram, Stanley (1974). *Obedience to authority : an experimental view*. London: Pinter & Martin, 1997.
- Mill, John Stuart (1987 [1871]). Utilitarianism. In A. Ryan (Ed.), *Utilitarianism and Other Essays*. Harmondsworth: Penguin Books, pp. 272-338.
- Minow, Martha (1998). *Between vengeance and forgiveness : facing history after genocide and mass violence*. Boston: Beacon Press.
- Nagin, Daniel S. (2000). Deterrence and Incapacitation. In M. Tonry (Ed.), *The Handbook of Crime and Punishment*. Oxford: Oxford University Press, pp. 345-368.

- Natsios, Andrew (2008, July 12, 2008). A Disaster in the Making. *ICC, Making Sense of Darfur*. Retrieved May 29, 2009, from <http://www.ssrc.org/blogs/darfur/2008/07/12/a-disaster-in-the-making/>
- OECD (2009). Net ODA from DAC countries from 1950 to 2008 (updated April 2009). Retrieved May 29, 2009, from <http://www.oecd.org/dataoecd/43/24/1894385.xls>
- Peirce, Charles Sanders (1868). Some Consequences of Four Incapacities. *Journal of Speculative Philosophy*, 2, 140-157.
- Plato (1901). *The republic of Plato; an ideal commonwealth (Translated by Benjamin Jowett)* (B. Jowett, Trans. Rev. ed.). New York,: The Colonial Press.
- Regulation (EC) No 1907/2006 of the European Parliament and of the Council, 1907/2006 Cong. Rec.(2006), from <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:396:0001:0849:EN:PDF>
- Reason, James T. (2000). Human error: models and management. *British Medical Journal*, 320(7237), 768-770.
- Roche, Declan (2005). Truth Commission Amnesties and the International Criminal Court. *The British Journal of Criminology*, 45(4), 565-581.
- Staub, Ervin (1989). *The roots of evil : the origins of genocide and other group violence*. Cambridge: Cambridge University Press.
- Strawson, P. F. (1987). *Skepticism and naturalism : some varieties : the Woodbridge lectures 1983*: London : Methuen, 1985 (1987 [printing]).
- Tallgren, Immi (2002). The Sensibility and Sense of International Criminal Law. *European Journal of International Law*, 13(3), 561-595.
- Trolle, Jørgen (1945). *Syv Maaneder uden Politi. 2. Opl*: Kjøbenhavn: Nyt Nordisk Forlag.
- Walker, Nigel (1991). *Why punish?* Oxford: Oxford University Press.
- Wilson, James Q. (1983). *Thinking about crime : Rev ed*. N.Y.: Basic Books.
- Wippman, David (1999). Atrocities, deterrence, and the limits of international justice. *Fordham International Law Journal*, 23, 473-488.
- Wright, Richard, W. (2001). Once More into the Bramble Bush : Duty, Causal Contribution, and the Extent of Legal Responsibility. *Vanderbilt Law Review*, 54(3), 1071-1132.