



FACULTY OF LAW
Lund University

Monika Bar

Who should bear the blame? The allocation of criminal liability for greater deterrence in international criminal law

JAMM04 Master Thesis

International Human Rights Law
30 higher education credits

Supervisor: Christoffer Wong

Term: Spring 2016

Contents

SUMMARY	3
PREFACE	5
ABBREVIATIONS	6
1. INTRODUCTION AND BACKGROUND	7
1.1 INTRODUCTORY COMMENTS	7
1.2 RESEARCH QUESTION AND METHODOLOGY	10
1.3 CURRENT STATE OF RESEARCH AND THE CONTRIBUTION OF THIS DISSERTATION	11
1.4 MATERIALS AND METHOD	16
1.5 STRUCTURE	18
2. THE CASE OF ERDEMOVIĆ	19
3. UNDERSTOOD ALLOCATION OF LIABILITY FOR CRIMES IN INTERNATIONAL LAW	23
3.1 INTRODUCTION TO THE STRUCTURE OF INTERNATIONAL CRIME	23
3.2 IMPOSITION OF LIABILITY	24
3.2.1 <i>MODES OF LIABILITY</i>	24
3.2.2 <i>INDIVIDUAL CRIMINAL RESPONSIBILITY</i>	26
3.2.3 <i>COMMAND RESPONSIBILITY</i>	27
3.3 RELEVANT DEFENCES EXCUSING LIABILITY	29
3.3.1 <i>OBEDIENCE TO SUPERIOR ORDERS</i>	29
3.3.2 <i>DURESS</i>	32
3.3.3 <i>SUPERIOR ORDERS AS AN ASPECT OF DURESS</i>	35
3.4 WHAT IS MISSING FROM THE LAW?	37
4. UNDERSTANDING HUMAN BEHAVIOR	40
4.1 THE LEGAL UNDERSTANDING OF HUMAN BEHAVIOR	40

4.2 THE REALITY OF HUMAN BEHAVIOR	44
4.2.1 CONFORMITY	45
4.2.2 OBEDIENCE	47
4.2.3 INSTITUTIONAL INFLUENCE	50
4.3 “THE TRAITOR IN YOUR SKULL”	52
4.3.1 THE REALITY OF DECISION MAKING	53
4.3.2 “MORAL CHOICE”	60
4.3.3 “REASONING AGENTS”	64
5. LAW AS A DETERRENT	67
5.1 THE LEGAL UNDERSTANDING OF DETERRENCE	67
5.2 THE REALITY OF DETERRENCE	68
5.3 ALLOCATING LIABILITY TO ACHIEVE GREATER DETERRENCE	69
5.4 POTENTIAL PROBLEMS WITH ALLOCATING LIABILITY AS PROPOSED	73
5.4.1 COGNITIVE ERROR IN LAW MAKING	73
5.4.2 WHAT WOULD CONSTITUTE EFFECTIVE TRAINING?	75
5.4.3 THE PHYSICAL PERPETRATORS	77
5.4.4 DEFINING THE GUILTY	78
6. CONCLUSION	80
BIBLIOGRAPHY	82
TABLE OF CASES	85

Summary

Deterrence of international crime is one of the key goals of international criminal law. To constitute effective deterrence, criminal liability must be allocated on the basis of an accurate understanding of what causes criminal behavior. And yet, the current allocation of criminal liability – primarily on the physical perpetrators of the crime and those directly involved in it – appears to be based on an antiquated model of human thought and behavior. The key elements of that model appear to be mental/physical dualism (the idea that the mind is somehow separate from the physical body/brain), and the notion of the mind as a unitary organ of consciousness – able to consciously control the actions of the body at all times, unless disabled by illness or injury. Neither of these ideas can be reconciled with the evidence emerging from behavioral research. In particular, the idea – inherent in much of the jurisprudence that considers liability for international crimes – that most of human action emerges from a primarily rational cognitive process of decision-making over which one has the power to exercise conscious control has been substantially undermined, if not entirely demolished.

The jurisprudence indicates that those who form and enforce the norms of ICL expect – in reliance on the flawed model of human behaviour described above – that the physical perpetrators are likely to know about the relevant legal prohibitions and/or are able to consider them and weigh them in the balance as part of a conscious decision making process that ultimately results in the crimes they commit. This expectation would seem unlikely to be fulfilled in view of the research on human behaviour, which indicates much of it to be the result of unconscious biases and habits, as well as social pressures, that the actor might not realize are operating upon him, let alone be able to control. Further, engaging in the sort of rational cognitive process that is able to weigh pros and cons of a decision and deliberately take legal principles into account, requires time, energy, and a lack of emotional

investment in the outcome – qualities which are generally in short supply in those engaged in a conflict. It would therefore follow that, if most behaviour – particularly behavior in a conflict - is not the result of a primarily-rational decision-making process able to take the law into account, the average combatant cannot realistically be deterred from committing a crime even if he has prior knowledge of the international criminal rules that prohibit the conduct he is embarking upon.

This dissertation will therefore examine the possibility that in order for the international criminal law to have a greater impact, it must target decision-makers higher up the chain of command from those who physically perpetrate international crime. Specifically, it will be suggested that those responsible for the training and socialization that a combatant receives should be presumed to bear responsibility for any international crimes that the combatant subsequently commits (unless that presumption is rebutted by evidence showing the training/socialization was not a significant factor in the crime in question). This is because those are the people who appear to be best placed to create the habits and pressures that will operate on the combatant in the midst of a conflict, and therefore have the power to affect – or prevent – the commission of international crime.

Preface

My fascination with psychology and related areas of behavioral science began with an introductory course I had taken in my first year of university. Although I ultimately decided to study law, I continued to pursue my interest in psychology alongside my legal career. As a result, I was often struck by the extent to which the law's expectations about human behavior can diverge from reality.

With the emergence of the field of legal psychology, some of that divergence is being addressed, at least in relation to domestic law. However, it seems that little is being done to correct the misconceptions about human behavior that plague international criminal law, despite the increased concerns about its apparent ineffectiveness in deterring the perpetrators of atrocities from committing their crimes. This thesis hopes to address one or two of these misconceptions, and to suggest a possible way to increase the deterrent effect of the law.

I was greatly aided in completion of this project by my supervisor, Christoffer Wong, who patiently guided me in the process of writing a thesis and encouraged the development of my ideas. I would also like to thank my classmates at Lund University, as well as my friend Ela, whose companionship and friendship helped to keep me sane through the long months of research and writing.

Finally, I would like to thank my mother for her steadfast support.

Monika Bar
27 September 2016
BVI

ABBREVIATIONS

ICC	International Criminal Court
ICL	International Criminal Law
ICTY	International Criminal Tribunal for the Former Yugoslavia
ICTR	International Criminal Tribunal for Rwanda
The Rome Statute	The Rome Statute of the International Criminal Court
The Genocide Convention	Convention on the Prevention and Punishment of the Crime of Genocide 1948

1. Introduction and Background

1.1 Introductory Comments

Legal scholars recognize that international criminal law tends to concern itself chiefly with “collective crimes” – crime committed by way of co-operation or collaboration of a large number of persons, acting within an established network of some kind (military, state or perhaps ethnic or tribal group) and working towards a common goal.¹ Where that goal is criminal in itself – for instance, where one of the warring groups is explicit in its genocidal intentions towards the enemy group – the courts have acknowledged the importance of the whole group to the commission of the crime, and sought to penalize those within it who took advantage of their positions of power to avoid direct physical involvement in any specific crimes, but rather planned, directed, and encouraged the criminal project along from the safety and comfort of their offices. The International Criminal Tribunal for the Former Yugoslavia, for instance, summarized the law’s position on international crimes as follows:

Most of these crimes do not result from the criminal propensity of single individuals, but constitute manifestations of collective criminality: the crimes are often carried out by groups of individuals acting in pursuance of a common criminal design. Although some members of the group may physically perpetrate the criminal act...the participation and contribution of the other members of the group is often vital in facilitating the commission of the offence in question. It follows that the moral gravity of such participation is often no less – or indeed no different – from that of those actually carrying out the acts in question.²

International criminal law’s chief concern thus far appears to have been how to connect the criminal to the crime, where the criminal in question is not the physical perpetrator. Various modes of liability – such as Joint Criminal Enterprise or Co-Perpetration – have been developed to

¹ See for instance, G. Werle (2007) “Individual Criminal Responsibility in Article 25 ICC Statute,” *Journal of International Criminal Justice* 5(4), pp. 953-976 at 953-954; and M. Osiel (2005) “The Banality of Good: Aligning Incentives Against Mass Atrocity,” *105 Columbia Law Review*, pp.1751-1862 at 1752 and 1753

² Judgment, *Tadić* (IT-94-1-A), Appeals Chamber, 15 July 1999 at para 191.

address this problem, together with different forms of accessory and inchoate liability for “support” members within the criminal group. However, where the group’s goal is not criminal in itself – where, for instance, in the course of a regular conflict over territory or even peacekeeping operations torture or murder of civilians is committed on a large scale – it becomes much harder for the law to understand the crimes committed within the context of the group, and of the conflict. Further, the subtle – yet powerful – influence that the group itself (criminal or not) has on the commission of international crime continues to be mostly unrecognized in international criminal jurisprudence. Because of this, the law cannot seriously address such questions as, why each horrible crime it deals with was committed, and what drove each perpetrator to commit it? More importantly, how can we stop similar crimes being committed in the future?

Considering the jurisprudence of international criminal law overall – that is, looking through a number of cases in which evidence of international crime, and the factors leading up to its commission is given – one can start to roughly classify the crimes committed by members of opposing groups in the course of an armed conflict into three general types. The first is opportunistic crime – crime committed simply because there is an opportunity to do so in the chaos, being either theft, rape or murder which the perpetrator initiates on his own volition. The second is crime committed by mistake or in the grip of insanity, under the influence of some confusion or mental illness that the law recognizes as robbing the perpetrator so completely of sufficient intent or control over his actions, that a just imposition of criminal liability is made impossible. The third type is crime committed under social and circumstantial pressure – most commonly, through obedience to superior orders of a recognized authority, duress or some other type of pressure known to operate in group situations, particularly in groups of a hierarchical nature.

The lines between the three cannot be sharply drawn – there is sometimes very little, if anything, to differentiate duress from insanity, or opportunistic crime from one motivated by the tribal hatreds inherent in war

and encouraged (implicitly or otherwise) by one's peers, superiors and society. Nevertheless, when considered broadly, the first two types identified above are not particularly interesting in the context of group conflict, given that they are generally committed by lone individuals. This dissertation will proceed on the assumption that, as such, they are not normally capable of generating the sort of mass atrocities that attract the interest of international criminal law.

It is therefore the third type of crime – that committed within, and under the influence of, one's group during a violent inter-group struggle – that deserves greater focus in ICL, and in this dissertation. However, for all intents and purposes, the pressures generated by a group engaged in a conflict – its causes and effects, its causal contribution to the crime - are all but ignored by ICL. The relevant law continues to target primarily those individuals whose contributions to the crime can be identified as intentional – the physical perpetrator, the person who ordered or otherwise planned the crime to be committed - focusing very narrowly on those individuals' causative contribution in the same way that domestic criminal law, which tends to deal with individual crimes committed during peacetime, does. Outside of the limited concepts of command liability and indirect perpetration, ICL ignores those further up the chain of command responsible for shaping the social forces that caused or contributed to the perpetrator's commission of the crime. Little thought seems to be given to whether, in light of the widespread devastation that international crimes are capable of, the old principles of individual criminal responsibility imported from domestic jurisprudence should not be reconsidered, and expanded to account for the causative impact of social pressure – without having to descend to some form of collective liability³ - in order to target more accurately those who may be truly in position to prevent atrocities from being committed.

³ Collective responsibility would not seek to attribute blame to, or punish, the individual or individuals that committed the crime, but rather the network – the state, military organization, etc – that they were part of. Usually, the penalty imposed consists of financial reparations to the victim. For example, Mark Osiel proposes an imposition of strict collective responsibility on senior military officers as a means of deterrence – see: Osiel (2005)

This dissertation will suggest that a different approach to crimes committed under the influence of the group is needed in order for international criminal law to provide a more effective deterrent to such crimes. Instead of focusing on the physical perpetrator of such crime (who, as the research examined below indicates, might be unable to resist the pressures acting upon him), this approach would require paying greater attention to the social and psychological pressures that induce the perpetrator to commit the crime, and allocating responsibility for the crime to those who are responsible for creating the relevant pressures and/or in the position to prevent or restrain their creation.

1.2 Research Question and Methodology

This dissertation will examine the extent to which the current allocation of criminal liability for international crime committed under social pressure is congruent with the aim of deterrence⁴ in international criminal law. For the purpose of this dissertation, it will be assumed that deterrence is or should be one of the key aims of ICL – perhaps even the primary aim. This assumption is considered to be both unproblematic and self-evident, particularly in light of numerous pronouncements emphasizing the importance of prevention of international crimes, including that contained in the Preamble to the Rome Statute of the International Criminal Court.⁵

The dissertation will progress through three aspects, answering the following questions:

- (1) What is the current allocation of criminal liability for international crime? The aim will be to clarify the legal basis for individual and command responsibility relevant to social-pressure related crime in particular, as well as the defences⁶ of superior orders and duress that excuse liability for such crime. The relevant rules will be identified by reference to the Rome Statute, as interpreted by

⁴ How deterrence is currently thought by judges to operate in ICL is examined below, at 5.1.

⁵ “The State Parties to this Statute...Determined to put an end to impunity for the perpetrators of these crimes and thus contribute to the prevention of such crimes...”

⁶ “Defences” in this context mean excuses from liability, rather than justifications.

the International Criminal Court where such interpretation is available, or where not available, by reference to pre-ICC jurisprudence recognized as concerning international crimes.

(2) What is the reasoning behind the current allocation, and is that reasoning likely to be well founded? The dissertation will seek to explore the understanding of human behavior that underpins the current law, and the assumptions concerning the impact (actual or potential) of the law in altering undesirable behavior. The results of behavioral research will then be presented as a counterpoint; setting out what the evidence is for how humans behave in reality, and whether it differs from the understanding that underpins the law.

(3) How effective is the current law as a deterrent, and what can be done to increase its effectiveness? The likelihood of the current law serving as an effective deterrent will be evaluated in light of evidence from behavioral research. An alternative attribution of liability will then be proposed, targeting the root causes of international crimes identified by relevant research, as a more likely prospect for having a deterrent effect.

The overarching question that this dissertation will seek to answer is how, in light of behavioral research findings relevant to causal factors in the commission of international crimes, should liability be allocated for the law to have a more reliably deterrent and/or preventive effect.

1.3 Current state of research and the contribution of this dissertation

This dissertation falls within an interdisciplinary field of inquiry, which seeks to apply research concerning human behavior – chiefly psychology and neuroscience – to legal concepts, norms, and procedures. This intersection of law and behavioral science is generally concentrated on the application of empirical research findings to a limited number of issues primarily relevant to domestic criminal and family law, and examining the

behavior of some of the actors within the legal system. Typical areas of interest within this field can be discerned by reference to one or more of its few introductory texts, such as Kapardis' *Psychology and the Law: A Critical Introduction*.⁷ Issues relating to testimony of eyewitnesses (both adult and children) appear of be of greatest concern – Kapardis dedicates four chapters to matters within this topic – encompassing matters such as the dubious reliability of memory, and the even more dubious reliability of eyewitness identification of perpetrators. Other major concerns appear to be jury decision-making, sentencing disparities, detecting deception (which, arguably, also concerns witnesses), the role of psychologists as expert witnesses, and the psychology of law enforcement officers.

Whilst this interdisciplinary area – referred to sometimes as legal psychology – is slowly starting to branch out beyond the ambit of domestic criminal law, research remains limited by a number of factors, including the small percentage of psychologists working in this field, the focus of research on each researcher's own country, the lack of knowledge of law and procedure by the psychologists who are interested in the field, as well as the corresponding lack of knowledge of psychology within the legal profession and the reluctance to consider behavioral science seriously as a result.⁸ It is therefore not surprising that there is very little of such interdisciplinary research available specific to the issues of concern in international criminal law, along with little examination of the assumptions about human behavior that arise from its jurisprudence and underlie its legal doctrines. Arguably, much of the research addressing issues in domestic criminal law may be readily transferrable into the international arena - such as general issues concerning reliability of eyewitness testimony, which may be the same regardless of the forum before which the witness appears. However, the issues that are more unique to ICL – problems of enforceability and legitimacy (not often encountered by domestic courts), as well as insights into the kind of perpetrators it deals with and the scale of their crimes – remain as yet largely unexplored by this field. Further,

⁷ A. Kapardis (2010) *Psychology and the Law: A Critical Introduction*, 3rd Ed, Cambridge: Cambridge University Press

⁸ *Ibid*

interdisciplinary research of this sort appears to have had little to no impact on the norms, procedures and scholarship of ICL – unlike in domestic criminal systems, where findings concerning, for instance, eyewitness identification and social pressure have had some success in improving police investigation and interrogation procedures.⁹

The particular issue of the effect of social pressure on international crimes has not previously been much considered by legal jurisprudence, although it has been discussed by scholars both within and outside the legal field. The discussion thus far has mostly revolved around whether punishment for a crime committed under the pressures inherent in a conflict can be morally justified. One major recent contribution is Natasha Gonzalez' dissertation (from the point of view of psychology) on the availability of excuses – superior orders or duress – to defendants facing charges of international crimes.¹⁰ Gonzalez explores much of the relevant psychological research in the context of military operations to critique the limitations of current defences and to argue for their greater contextualization – for the courts to take into account the effects that, *inter alia*, the stresses of combat and military culture have on the commission of the crime. Gonzalez's focus is to make proceedings more fair and just for defendants accused of crimes they committed under pressure and to “highlight the power of [...] situational forces and constraints that speak volumes in terms of culpability.”¹¹ However, Gonzalez only briefly considers the impact that allowing the court to take into account the circumstances under which the crime was committed would have on the administration of justice for international crimes, and her conclusions in this respect are unconvincing. She acknowledges concerns that expanding the court's consideration in this way would lead to an inability to convict virtually anyone, which were previously expressed in the context of defendants raising social deprivation or background abuse to excuse crimes

⁹ See the discussion in T. Gilovich et al (2016), *Social Psychology, 4th Ed*, New York and London: W. W. Norton and Company at pp.600-618

¹⁰ N. Gonzalez (2004), *Moral Monsters or Ordinary Men Who Do Monstrous Things? Psychological Dimension of the Military and Their Implications for War Crimes Tribunal Defences*, unpublished PhD thesis, Widener University.

¹¹ *Ibid*, p.84

they were charged with in domestic proceedings.¹² However, she dismisses such “slippery slope” concerns on the basis that such novel defences rarely lead to an acquittal, so that most defendants could not avail themselves of a contextual defence that she proposes.¹³

This seems somewhat illogical, given that the situational forces (social and circumstantial pressures) that she describes can be said to apply to almost any soldier or other combatant who commits a crime in the course of a group conflict. The fact that similar defences in domestic settings have been rejected in the past is no answer to this difficulty, not least because she is advocating for the law to expand – the rejection of contextual defences and the lack of acquittals in a system that does not take the broader context into account cannot be used as evidence of what would happen in a system, such as the one she proposes, that would take such broader context into account. Therefore, if the only change that is made is to allow defendants to raise contextual defences – forcing the judges to take into consideration the whole gamut of exculpating pressures under which the defendant committed the crime – it is unlikely that ICL would be able to convict most of those who appear as defendants in its system.

Ziv Bohrer, a legal scholar, agrees that this would be the case if such defences were available.¹⁴ He has attempted to answer Gonzalez’s criticism of the limits imposed on available defences, as well as the criticism – raised, for instance, by Mark Osiel¹⁵ – that in light of the effects of social and circumstantial pressures known to operate in times of conflict, prosecutions in ICL are unjustifiable. He maintains that such prosecutions (of defendants who committed their crimes under what he calls “sociopsychological coercion”) can be morally justified because the ICL’s system of laws and law enforcement forms a normative assertion that “can affect an individual’s disposition, thereby increasing the ability of a person to rationally decide

¹² *Ibid*, pp.77-79

¹³ *Ibid*, pp.88-89

¹⁴ Z. Bohrer (2012), “Is the Prosecution of War Crimes Just and Effective? Rethinking the Lessons from Sociology and Psychology, *Michigan Journal of International Law*, Vol.33, pp.749- 819 at p.772

¹⁵ See, for instance: M. Osiel (2001), *Mass Atrocity, Ordinary Evil and Hanna Arendt: Criminal Consciousness in Argentina’s Dirty War*, Yale University Press quoted in Bohrer (2012) at p.773; and Osiel (2005)

whether or not to commit the prohibited act.”¹⁶ He relies on results of certain psychological research, showing such an effect in respect of domestic criminal law, in support of this assertion. However, he concedes that certain individuals who had suffered from an extreme form of socio-psychological conditioning – such as brainwashing – prior to committing a crime, should be afforded a defence on that basis.

Bohrer’s work is disappointing in one respect, and that is the apparent selective use of research findings. He seems to cherry-pick a few findings on the effect of legal norms on decisions in the domestic setting, mostly ignoring the vast body of research on decision-making that puts those findings in context. He confidently assert that the “mainstream” view of psychology finds that a spectrum exists between fully rational and irrational behavior - implying that this spectrum is comparable to that acknowledged in criminal law. He therefore concludes that “[r]ational, dissuadable behavior can still be assumed to be the default mode of human behavior.”¹⁷ This is simply misleading; whilst it is true that both criminal law and psychology acknowledge a certain spectrum of rationality, they are poles apart in their opinions on where most people in most situations fall on that spectrum - with the body of psychological research (explored in more detail below) suggesting that it is mostly irrational decisions¹⁸ that are the norm and criminal law assuming the opposite.¹⁹ Given that his entire argument rests upon a contentious assumption – that rational behavior is the norm – the fact that he fails to examine it in any detail undermines the rest of his work; as does his apparent assumption, evident throughout, that every “good” decision is the product of a rational process.

¹⁶ Bohrer (2012) p.754

¹⁷ *Ibid*, p.762

¹⁸ “Irrational” decisions are meant here to mean those decisions that motivated (or caused) more by heuristics, biases, and social and circumstantial triggers than conscious logical reasoning. The current position of mainstream cognitive psychology appears to be that human thinking can only be called “rational” in most cases if we redefine “rationality” to mean an effective use of heuristics, and forget about logic. For a summary of current discussion, see M. W. Eysenck and M. T. Keane (2015) *Cognitive Psychology: A Student’s Handbook, 7th Ed*, London and New York: Psychology Press at pp.623-628.

¹⁹ Although to be fair, it is not entirely clear what Bohrer means when he speaks of “rational, dissuadable behavior,” I assume it means behavior based on a decision that is closer to full rationality (logical reasoning) than irrationality on the spectrum.

This dissertation will attempt to take this discussion a step further; beyond considering *whether* we should punish – whether more contextual defences ought or ought not be available, and how to justify punishment – and consider instead *who* we should punish. If those who physically perpetrate international crime cannot be reliably deterred by the law (and it will be argued that, for the most part, it is unrealistic to expect the law to influence their behavior), then the law must target those who have the power to influence the commission of the crime and can be deterred – and this dissertation will therefore attempt to explore the question of where liability for international crime ought to be shifted to achieve more reliable deterrence, in light of relevant empirical findings. In this, it is proposed to more broaden the scope of inquiry into behavioral research beyond simply considering how an individual is affected by the group, to also consider why he is so affected - what science can tell us about how decisions are made in general. In order to get away from the moral justification of punishment argument, it will be assumed for the purpose of this dissertation that punishment itself can be justified, and it is unimportant here whether this justification can be made on a consequentialist or retributive basis.

1.4 Materials and method

Whilst ICL has several sources, including customary law and the jurisprudence of various military and international tribunals, this dissertation will use the Rome Statute, as interpreted by the ICC, as the primary source for relevant contemporary rules of international criminal law. This is because the ICC appears to be the most authoritative source of relevant jurisprudence, having jurisdiction over international crimes committed on the territory and/or by nationals of 124 countries party thereto,²⁰ and able to exercise jurisdiction even over non-State Parties upon referral from the Security Council of the United Nations. Where interpretation by the ICC is yet unavailable, the interpretation of relevant legal concepts will be based on the most authoritative modern jurisprudence and academic commentary

²⁰ Articles 12 and 13 of the Rome Statute set out the court's jurisdiction; Article 17 limits that jurisdiction to cases where the State with jurisdiction over a particular matter is either unwilling or unable to carry out a genuine investigation and/or prosecution.

available. Standard legal method will be employed in the discussion of the relevant rules.

Case law will also be used as a source of judicial pronouncements and commentary on the subject of human behavior and its causes. The judgments employed in this context should not be considered precedent or authority for a legal rule, but rather as illumination of the understanding behind the rule – evidence of how judges think about and understand concepts associated with behavior, such as moral choice or rationality, on the basis of which they then go on to develop and apply the law. Further, the case of *Erdemović*, discussed in more detail below, will be used to ground the discussion in concrete facts of a specific case and provide for a real-life example of the issues explored. Whilst *Erdemović* predates the ICC (being a matter before the ICTY), it remains the best modern example of how the courts deal with the complexities of crime that is the result of powerful social pressures arising during an inter-group conflict.

Finally, the dissertation will utilize findings of empirical research conducted in the field of behavioral science – primarily results of psychological studies and experiments, but also occasionally dipping into the field of neuroscience – to examine the validity of the understandings and assumptions about human behavior upon which relevant law is based, and to form the basis of the alternative attribution of liability being proposed. Because the results of any “cutting edge” research can often be subject to some criticism and controversy, the research selected will be that which is broadly, though perhaps not universally, accepted within its field. This would comprise findings that have been successfully replicated and/or fit within a larger body of similar experiments, preferably to the point where they can be included in an introductory text for first-year students of the subject, supported by explanatory material from acknowledged experts in this field. The material included is intended to highlight the most relevant findings, whilst situating them in within the broader context of the research on behavior, judgments and decisions.

1.5 Structure

Chapter 2 will set out the facts in and discuss the case of *Erdemović*, in order to ground the discussion in a concrete modern example of how international criminal law treats crimes committed under group and circumstantial pressure. The discussion in Chapter 3 will move from there to elaborate on the rules that determine the allocation of criminal liability at present, setting out the rules governing individual and command responsibility, together with the two relevant defences – obedience to superior orders and duress. Chapter 4 will explore the reasoning behind the present rules – the legal understanding of human behavior upon which the rules are based – before moving on to set out relevant evidence from behavioral science, and analyze the legal understanding in light of that evidence to determine. This would be done to determine whether the allocation of criminal liability, based on that understanding, is likely to have any deterrent effect upon those it aims to deter. This will in turn lead into the discussion of the proposed allocation of criminal liability in ICL in Chapter 5, including any problems that may arise in relation to that allocation, before coming to a brief conclusion in Chapter 6.

2. The case of Erdemović

On 16 July 1995, on the grounds of a collective farm near Pilica in the former Yugoslavia, a 23-year-old soldier called Dražen Erdemović participated, as a member of an execution squad, in a massacre of Bosnian Muslim civilian men from Srebrenica who had surrendered to the Bosnian Serb forces. Around 1,200 unarmed men and teenage boys died that day; by his own estimate, Erdemović personally shot and killed about 70 of them.²¹

On his account – accepted by both the prosecution and judges of the International Criminal Tribunal for the Former Yugoslavia, before which he was eventually tried and convicted for a crime against humanity – Erdemović was not a willing participant in this slaughter. Prior to his arrival at the farm, he was not aware of what he would be required to do, and then it took a direct threat from his superior – combined with the lack of any overt support from the seven other soldiers in his unit, who co-operated without demur - to make him carry out the order to kill following his initial resistance.²² Entering a guilty plea to the charges put to him, he added the following statement – later considered by the court to amount to an assertion of the defences of obedience to superior orders and/or duress:²³

Your Honour, I had to do this. If I had refused, I would have been killed together with the victims. When I refused, they told me: 'If you are sorry for them, stand up, line up with them and we will kill you too. I am not sorry for myself but for my family, my wife and son who then had nine months, and I could not refuse because they would have killed me.'²⁴

It seems clear from the facts accepted by the court that Erdemović was subsequently haunted by the fact that he gave in, and followed the order given. His guilt, constantly reiterated before the court, was palpable – he

²¹ The facts of the case, unless stated otherwise, are taken from the version of events set out by the court in Sentencing judgment, *Erdemović* (IT-96-22-T), Trial Chamber, 29 November 1996, at paras 76-82.

²² According to Erdemović, he was only able to successfully resist committing another atrocity he was ordered to commit, because he had the support of three other members of his unit: *Ibid*, para 81.

²³ Sentencing judgment, *Erdemović* (IT-96-22-T), Trial Chamber, 29 November 1996, at para 14

²⁴ Sentencing Judgment, *Erdemović* (IT-96-22-Tbis), Trial Chamber, 5 March 1998, at para 14

confessed the particulars of his participation in the crime to a journalist (his involvement in any crime was not known to any investigating authority prior to him coming forward²⁵), confessed to the court at first instance and confessed again to the prosecutor immediately upon his arrival at the Hague. He co-operated with the Office of the Prosecutor in a way that was, as the Court noted in mitigation, “voluntary and unconditional” as well as “substantial, full and comprehensive,” including divulging serious crimes committed by Bosnian Serb forces that the prosecution had not previously been aware of.²⁶ He also suffered from post-traumatic stress disorder of such severity that the expert report prepared for the ICTY concluded in June 1996 that he was “insufficiently able to stand trial.”²⁷

From the background facts, as well as the above clear and credible expressions of guilt and remorse, it might be surmised that Erdemović was overall a man with a strong and well-developed conscience. That much is evident by the fact that, alone of the men in his unit, he initially refused to follow his commander’s orders – standing up not only to the authority of his legitimate superior, but also against the pressure of his peers’ compliance. According to his testimony, he only joined the Bosnian Serb army because his attempts to avoid the war – and escape to Switzerland – had been unsuccessful; he was far from an enthusiastic participant, but needed some security for himself and his family as a Croat in a Serbian republic. During the course of his service, he saved the life of an acquaintance who had been apprehended by Serbian forces (the man appeared before the Trial Chamber as his witness), and was demoted for refusing to carry out a mission likely to cause civilian losses.²⁸ There is no hint on the facts of any nationalistic fervor, hatred of Muslims, or anything else that would suggest a willingness to commit mass murder – in fact, one of his witnesses described him as part of a “multi-ethnic group of friends” and a person who “hated the war” and “was not a nationalist.”²⁹ If these facts about him are accepted, then the

²⁵ *Ibid*, para 16(i)

²⁶ Sentencing judgment, *Erdemović* (IT-96-22-T), Trial Chamber, 29 November 1996 paras 96-99

²⁷ *Ibid* para 5

²⁸ *Ibid*, paras 79 and 107.

²⁹ *Ibid* para 108.

picture that emerges is of a man whose participation in the horrific crime he was convicted of seems puzzling to say the least, and deserving of an explanation - even when the alleged coercion is taken into account.

And yet, in the legal analysis of Erdemović culpability for the crime no such explanation is forthcoming. As far as the judges were concerned, Erdemović had the possibility of exercising, and did exercise, a “moral choice” to commit the crime in question.³⁰ That choice was whether to obey and save his own life (and possibly that of his family) or disobey, as was his duty, and avoid committing mass murder – he chose wrongly, and so he was culpable.³¹ But if this apparently decent man can commit such a crime, would anyone in his situation really be able to make a different choice? And how realistic is the law’s expectation that they should?

In their assumption of the possibility of “moral choice” in this context of a subordinate complying with the orders of his superior, the judges were forced to engage in a realm of expertise that they were entirely unfamiliar with, and yet frequently called upon to consider – the study of human behavior and its causes in group conflict situations. Thus, while chiefly known for the elaboration of the law on duress and the validity of a guilty plea when accompanied by an apparently exculpatory statement,³² Erdemović’s case serves also as a notable illustration of how international tribunals deal with determinations of culpability in situations where the person who physically carried out the crime was clearly not the directing mind behind it – such as when the crime was committed following an order from a recognized authority, which the defendant felt unable to disobey. The understanding of human behavior behind such decisions is important because it informs not only decisions in particular cases, but also the development of the criminal law as a whole in the international sphere, determining how and upon whom the courts place primary criminal responsibility for some of the most heinous crimes known to man. This placement in turn affects the ability of criminal law to constitute an effective

³⁰ *Ibid* paras 18 and 19;

³¹ That this choice was difficult, and perhaps deserving of exculpation was acknowledged in the minority opinions of Judge Cassese and Judge Stephen in the Appeals Chamber.

³² See for instance, the discussion in T. Weingend (2012) “Kill or Be Killed: Another Look at *Erdemović*,” 10 *J. Int’l Crim. Just.* 1219-1237.

deterrent to, and end impunity for, such crime – the ability that it is, arguably, lacking at present.

3. Understood allocation of liability for crimes in international law

The contemporary international criminal law concerning liability for crimes – including those committed under social pressure - can be found mainly in the rules governing individual responsibility, crimes committed under the orders of a legitimate superior or under duress, and those governing the responsibility of a commander for crimes committed by his or her subordinates. Whilst such rules are found in a variety of sources, including customary law and the jurisprudence of military tribunals, the following discussion will set out the current law as administered and interpreted by arguably the most authoritative source of modern ICL – the International Criminal Court – noting its interpretation as well as any significant departures from, other sources of international jurisprudence, and touching briefly upon some relevant concerns.

3.1 Introduction to the structure of international crime

The Rome Statute lists four crimes to be within the jurisdiction of the ICC, three of which are of relevance to this discussion: genocide, crimes against humanity, and war crimes (Article 5(1)(a)-(c)).³³ Each of those crimes is in turn further defined at Articles 6 through 8, which set out the material elements of each crime, listing the prohibited conduct (*actus reus*) as well as the contextual factors that must accompany such conduct in order for an international crime to be committed. Much of the content is derived from earlier treaties – for instance, the definition of genocide at Article 6 is taken verbatim from Article II of the Genocide Convention – ensuring that

³³ The fourth crime – the crime of aggression – is defined as the planning, preparation, initiation or execution by a person in a leadership position of an act of aggression, where that act is a manifest violation of the Charter of the United Nations (Article 8 *bis*, adopted by way of an amendment to the Rome Statute at Kampala in 2010). The ICC will not be able to exercise its jurisdiction over this crime until after 1 January 2017. Further, given that this thesis aims to argue for a greater liability on leaders, there is no need to consider a crime that already can only be committed by leaders.

the Rome Statute, for the most part, remains consistent with the established norms of international criminal law.

The mental element (*mens rea*) necessary for the imposition of liability is defined in Article 30 to constitute intention and knowledge, "unless otherwise provided." The Rome Statute thus allows for a lesser degree of fault, such as for instance negligence in the case of command responsibility,³⁴ to suffice in certain circumstances. It should be noted that, unlike in most domestic criminal law, having the general criminal *mens rea* in relation to the prohibited conduct (the *actus reus*) does not suffice for an international crime to be committed. It must also be accompanied by the specific *mens rea* and/or contextual elements relevant to each crime.

Thus, for instance, while killing a person (*actus reus*) with the intention of doing so (*mens rea*) would be sufficient to form a crime within the jurisdiction of a domestic criminal court (murder), it would not suffice to generate liability within the jurisdiction of the ICC. In order to fall within that jurisdiction, additional elements would have to be shown. For genocide, the killing would have to be accompanied by the additional intent to "destroy, in whole or in part, a national, ethnical, racial or religious group" to which the victim belonged. For a crime against humanity (Article 7), the murder would have to be committed "as part of a widespread or systematic attack directed against any civilian population" and the perpetrator would have to have the additional mental element of the knowledge of that attack. For a war crime (Article 8), the killing would have to be committed within the context of an armed conflict, and the victim would have to be a person protected under the Geneva Conventions.

3.2 Imposition of liability

3.2.1 Modes of liability

Given that international crimes tend to be carried out by a group, the law must also provide for the various ways in which participation in and contribution to the crime is possible. Here too, there is a theoretical difference between domestic and international criminal law in how crimes

³⁴ See discussion below at 3.2.3

committed by more than one perpetrator are dealt with. Domestic systems tend to adopt either a unitary or a differentiated model³⁵ for how they deal with the individual roles of participants in a crime, and each participant is charged and punished depending on the degree of their participation in the crime. In a unitary model, all primary participants – those who were present at the scene of the crime and/or directly involved – will be charged with the same crime (for instance, a robbery), although their punishment might in practice be different and those who did not directly participate might be charged with some lesser offence connected to the main crime (ie, being an accessory after the fact). In a differentiated system, each participant will be charged with greater or lesser offences depending on their role in relation to the crime – the charges facing the person who committed the robbery will be greater than those of the person who drove the getaway car.

Under the ICC statute, each perpetrator is charged with a crime depending on a “mode of liability” in each individual case, prompting the court to classify each individual’s action into discrete categories of criminal activity rather than either charging all participants with the same offence or attempting to establish their relation to the crime committed. Each defendant is thus effectively a “principal,” and there is no automatic sentencing distinction, such as that which occurs in domestic criminal law between principals and accomplices.³⁶ The different modes of liability are enumerated under Article 25(3), attempting to account for every way that one could intentionally or knowingly commit or contribute to the commission of one of the crimes within the jurisdiction of the court – from physically carrying out the crime, through ordering and soliciting the crime to be committed, to contributing to the crime as part of a group with a common criminal purpose (joint criminal enterprise or co-perpetration). In addition, Article 28 sets out an additional mode of liability – command

³⁵ Description of the difference adopted from A. Cassese and P. Gaeta (2013) *International Criminal Law*, 3rd Ed, Oxford: Oxford University Press, at p.161-163

³⁶ *Ibid*

responsibility – though there is some debate over whether it can be properly classified as a "mode of liability" or a separate offence.³⁷

3.2.2 Individual criminal responsibility

The basis for individual criminal liability is thus set out in Article 25 of the Rome Statute as follows:

1. The Court shall have jurisdiction over natural persons pursuant to this Statute.
2. A person who commits a crime within the jurisdiction of the Court shall be individually responsible and liable for punishment in accordance with this Statute.
3. In accordance with this Statute, a person shall be criminally responsible and liable for punishment for a crime within the jurisdiction of the Court if that person:
 - (a) Commits such a crime, whether as an individual, jointly with another or through another person, regardless of whether that other person is criminally responsible;
 - (b) Orders, solicits or induces the commission of such a crime which in fact occurs or is attempted;
 - (c) For the purpose of facilitating the commission of such a crime, aids, abets or otherwise assists in its commission or its attempted commission, including providing the means for its commission;
 - (d) In any other way contributes to the commission or attempted commission of such a crime by a group of persons acting with a common purpose. Such contribution shall be intentional and shall either:
 - (i) Be made with the aim of furthering the criminal activity or criminal purpose of the group, where such activity or purpose involves the commission of a crime within the jurisdiction of the Court; or
 - (ii) Be made in the knowledge of the intention of the group to commit the crime;
 - (e) In respect of the crime of genocide, directly and publicly incites others to commit genocide;
 - (f) Attempts to commit such a crime by taking action that commences its execution by means of a substantial step, but the crime does not occur because of circumstances independent of the person's intentions. However, a person who abandons the effort to commit the crime or otherwise prevents the completion of the crime shall not be liable for punishment under this Statute for the attempt to commit that

³⁷ The distinction is not relevant for the purposes of this thesis, but for issues raised by the discussion, see for instance, M. Chantal (2007) "Command Responsibility: Mode of Liability for the Crimes of Subordinates or Separate Offence of the Superior," *J Int Criminal Justice*, Vol. 5(3), pp.619-637.

crime if that person completely and voluntarily gave up the criminal purpose.

Whilst liability under this Article is generally concerned only with the acts carried out by the individual himself, Article 25(3)(a) also provides for one mode of superior responsibility: indirect perpetration, where the crime can be committed “through another person.” In international law, this liability can be imposed upon a leader (or leaders) of a hierarchical organization, who plan the crime in question and execute it through their subordinates, using them as little more than tools. The control that the leader exerts over his subordinates must be such that compliance with his will becomes automatic. As an illustration, the ICTR Appeals Chamber in *Seromba* found that what is important was not that the defendant, Athanese Seromba, personally drove the bulldozer that destroyed the church, killing the refugees inside, but rather that he “fully exercised his influence over the bulldozer driver who, as the Trial Chamber’s findings demonstrate, accepted Athanese Seromba as the only authority, and whose directions he followed.”³⁸

3.2.3 Command responsibility

Article 28 of the Rome Statute imposes a liability upon those higher up in the chain of command for the crimes committed by their subordinates:

(a) A military commander or person effectively acting as a military commander shall be criminally responsible for crimes within the jurisdiction of the Court committed by forces under his or her effective command and control, or effective authority and control as the case may be, as a result of his or her failure to exercise control properly over such forces, where:

- (i) That military commander or person either knew or, owing to the circumstances at the time, should have known that the forces were committing or about to commit such crimes; and
- (ii) That military commander or person failed to take all necessary and reasonable measures within his or her power to prevent or repress their commission or to submit the matter to the competent authorities for investigation and prosecution.

³⁸ See discussion in Cassese and Gaeta (2013) at pp.178-179.

(b) With respect to superior and subordinate relationships not described in paragraph (a), a superior shall be criminally responsible for crimes within the jurisdiction of the Court committed by subordinates under his or her effective authority and control, as a result of his or her failure to exercise control properly over such subordinates, where:

- (i) The superior either knew, or consciously disregarded information which clearly indicated, that the subordinates were committing or about to commit such crimes;
- (ii) The crimes concerned activities that were within the effective responsibility and control of the superior; and
- (iii) The superior failed to take all necessary and reasonable measures within his or her power to prevent or repress their commission or to submit the matter to the competent authorities for investigation and prosecution.

The liability is therefore limited to the commanders with either actual or constructive knowledge of the specific crimes that are being committed (or are about to be committed) by those under their command. What actual knowledge amounts to in contemporary ICL was recently summarized in the ICC judgment in the case of *Bemba*.³⁹ The court held that actual knowledge cannot be presumed, but must be established by either direct or circumstantial evidence - it can be inferred from the circumstance, but that inference must relate to the knowledge of the accused rather than the general public or others in his organization. There is no need for the commander to have known every detail of the crimes committed, only to be aware that his troops' conduct qualifies as a war crime or a crime against humanity.⁴⁰

The constructive knowledge element under Article 28 differs slightly between military commanders and other superiors. Civilian leaders are only held responsible when they consciously disregard information that clearly indicates their subordinates are engaging in criminal activity, whereas military leaders are liable under a more all-encompassing standard of "should have known." Cassese believes that the civilian "consciously

³⁹ Judgment, *Prosecutor v Jean-Pierre Bemba Gombo*, (ICC-01/05-01/08), Trial Chamber III, 21 March 2016. Bemba, the former Vice President of the Democratic Republic of the Congo and leader of a militia known as Mouvement de Libération du Congo ("MLC"), was convicted on an number of counts for the murders, rapes and repeated pillage committed by those under his command.

⁴⁰ *Ibid*, paras 191-195.

disregard information” standard resembles that developed by the ICTY and the ICTR’s “had reason to know” formulation.⁴¹ This standard of constructive knowledge encompasses failure by a commander (whether military or civilian) to take steps to determine whether his subordinates are engaging in criminal activity, if he has information that puts him on notice of the risk of such activity. The information that constitutes the requisite notice “does not need to provide specific information about unlawful acts committed or about to be committed,”⁴² and the risk in question also does not need to amount to a clear possibility, but rather be merely sufficient to justify further inquiry.⁴³ As for constructive knowledge of military leaders, the ICC in *Bemba* made findings of actual knowledge, so did not consider it necessary to look at the “should have known” standard. However, the Pre-Trial Chamber in the same case considered it to be a standard amounting to negligence, imposing an “active duty on the part of the superior to take the necessary measures to secure knowledge of the conduct of his troops and to inquire, regardless of the availability of information at the time on the commission of the crime.”⁴⁴

Further, it is worth noting here that command responsibility embraces liability for omissions – failure to take action – as well as active commission of wrongful acts. This sets it apart from liability in indirect perpetration, which can only be imposed for active perpetration.

3.3 Relevant defences excusing liability

3.3.1 Obedience to superior orders

Article 33 of the Rome Statute allows for only a very limited excuse from liability when one’s crime is committed under the authority of one’s superior:

1. The fact that a crime within the jurisdiction of the Court has been committed by a person pursuant to an order of a

⁴¹ See discussion in Cassese and Gaeta (2013) at pp.189-190.

⁴² Judgment, *Delalić and others*, (IT-96-21-A), Appeals Chamber, 20 February 2001, at para 238

⁴³ Judgment, *Strugar*, (IT-01-42-A), Appeals Chamber, 17 July 2008 at paras 303-304

⁴⁴ Decision on the Charges, *Prosecutor v Jean-Pierre Bemba Gombo*, (ICC-01/05-01/08), Pre-Trial Chamber II, 15 June 2009 at paras 429 and 433

Government or of a superior, whether military or civilian, shall not relieve that person of criminal responsibility unless:

- (a) The person was under a legal obligation to obey orders of the Government or the superior in question;
- (b) The person did not know that the order was unlawful; and
- (c) The order was not manifestly unlawful.

2. For the purposes of this article, orders to commit genocide or crimes against humanity are manifestly unlawful.

The possibility of successfully invoking superior orders as a defence before the ICC seems therefore slim at best – given its limited resources, the ICC tends to concern itself mainly with perpetrators accused of committing the sort of acts one would struggle to ever mistake for being anything other than manifestly unlawful. Yet even this limited defence aroused some controversy, as it was seen by some as a dangerous departure⁴⁵ from the complete denial of superior orders as an excuse to liability – the “absolute liability” approach – established by the jurisprudence of past international criminal tribunals, and – it was argued - customary law.⁴⁶

This is perhaps not entirely a fair charge, as international criminal law has hardly been as monolithic in its adoption of the absolute liability approach prior to the Rome Statute as the above criticism might suggest. It is true that major international tribunals established before the ICC was created were explicit in their rejection of superior orders as excusing liability. For instance, Article 8 of the London Agreement, the treaty that established the Nuremberg Tribunal to deal with the crimes committed by the high-ranking Nazis, clearly set out that acting in pursuit of superior orders would not free a defendant from responsibility, though it may be considered as a mitigating circumstance. Similarly worded principles of absolute liability, with obedience to orders being seen as a mitigating factor

⁴⁵ See, for instance, the analysis in P. Gaeta (1999) “The defence of superior orders: The Statute of the International Criminal Court *versus* customary international law”, *European Journal of International Law*, Vol. 10, pp. 172-191

⁴⁶ *Ibid.* The approach adopted in Article 33 is similar to that traditionally seen in domestic tribunals, where superior orders can be a complete defence to liability unless the order is manifestly illegal or the perpetrator knew it was illegal (known as the “conditional liability” approach). Gaeta provides examples of states that ostensibly adopted this approach at pp.176-177, though he goes on to argue that the divergence between the two approaches in international practice (prior to Article 33) were more apparent than real, thus making for a customary law that adopts the absolutist approach.

at best, were subsequently adopted into the Charter of the International Military Tribunal for the Far East (at Article 6), as well as the Statute of the ICTY (Article 7(4)⁴⁷) and the Statute of the ICTR (Article 6(4), the operative terms of which are identical to that of ICTY). However, analysis of the relevant case law reveals that consistent application of this approach in ICL was always somewhat lacking, as the various judges of various tribunals wrestled with their sympathy for defendants caught between their duty to obey their superiors and the duty to obey the law. The jurisprudence of the ICTY Appeals Chamber in *Erdemović* provides one example of this conflict in the separate opinion of Judge Cassese, President of the Tribunal at the time. Unlike the majority, who agreed that superior orders did not amount to a defence *per se*, Judge Cassese seemingly rejected the absolute liability interpretation of Article 7(4) of the Statute of the ICTY. He argued that a soldier only has a duty to disobey an order that is either manifestly unlawful or known to the soldier to be unlawful, and may resort to a plea of superior orders when the order he followed was not obviously illegal⁴⁸ – thus adopting a position similar to that in Article 33 of the Rome Statute.

Examples of the conditional liability approach are also evident in the jurisprudence of earlier tribunals that considered international crimes. While the Nuremberg Tribunal, in accordance with the plain wording of Article 8 of the London Agreement, decisively rejected superior orders as ever amounting to a defence at all,⁴⁹ other post-WWII tribunals established by the Allies accepted such pleas as valid, at least theoretically. For instance, the Judge Advocate in the *Peleus Case*, acknowledged that it may not be fair to hold a subordinate responsible for an act, the lawfulness of which could only be determined upon a “careful consideration of questions of International Law”⁵⁰ – in other words, an act that was not obviously

⁴⁷ “The fact that the accused person acted in pursuant to an order of a Government or of a superior shall not relieve him of criminal responsibility, but may be considered in mitigation of punishment if the International Tribunal determines that justice so requires.”

⁴⁸ Separate and Dissenting Opinion of Judge Cassese, *Prosecutor v Dražen Erdemović* (IT-96-22A), Appeals Chamber, 7 October 1997 at p.15.

⁴⁹ See discussion in L. C. Green (1976) *Superior orders in national and international law*, Leyden: A. W. Sijthoff, at 178

⁵⁰ *Ibid*, at p.285. The case was tried under the British Warrant (Royal Warrant, 14 June 1945, Army Order 81/1945).

illegal. Other cases tried under the British Warrant adopted a similar approach, as did the US Military Tribunal.⁵¹ In *Re Buck*, for instance, the Judge Advocate accepted that a soldier's knowledge of the law concerning the rights of prisoners of war might be limited, and stated that a man would only be guilty if the order in pursuance of which he committed his crime was obviously unlawful or the accused knew it was unlawful.⁵² In *List and Others (Hostages Trial)* ("the *Hostages* case") the USMT commented that "if the illegality of the order was not known to the inferior and he could not reasonably have been expected to know of its illegality, no wrongful intent necessary to the commission of a crime exists and the inferior will be protected."⁵³

3.3.2 Duress

Under Article 31(1)(d) of the Rome Statute, duress is ostensibly a complete defence excusing the defendant from criminal responsibility:

1. (...) a person shall not be criminally responsible if, at the time of that person's conduct...

(d) The conduct which is alleged to constitute a crime within the jurisdiction of the Court has been caused by duress resulting from a threat of imminent death or of continuing or imminent serious bodily harm against that person or another person, and the person acts necessarily and reasonably to avoid this threat, provided that the person does not intend to cause a greater harm than the one sought to be avoided. Such a threat may either be:

- (i) Made by other persons; or
- (ii) Constituted by other circumstances beyond that person's control.

In order to satisfy the "duress" part of the above definition, the defendant must show that his crime "was the product of psychological coercion" and "his or her capacity to choose is so impaired as to require

⁵¹ *Ibid*, pp.286-288

⁵² *Ibid*, p286-287; *Re Buck* (1946) 13 Ann. Dig. 293

⁵³ *In re List and Others (Hostages Trial)*, Judgment 215 (19 February 1948)(USMT, Nuremberg), 15 ILR 632 at 650

excusing their liability.”⁵⁴ However, this leads to something of a logical contradiction with the requirement that the harm he causes is no greater than the harm he is seeking to avoid. On the one hand, what is required to establish duress is for the defendant to show that, in the circumstances, his autonomy was all but eliminated⁵⁵ “by asking him to sacrifice something (such as his family or himself) that he was not capable of doing”⁵⁶ - an absence of choice and an inability to do anything other than what the defendant was being forced to do. And yet, at the same time, the defendant is expected to make a choice – to evaluate the harm that he would be causing as a result of that duress, decide whether it is graver than the harm that he is seeking to avoid, and then choose whether or not he should comply. This rather assumes that a choice is possible – in which case, the component of “psychological coercion” cannot be fulfilled in the first place, leading the law to very much resemble the paradoxical requirements of Catch-22.

Aside from this, there is another major (and more often discussed) problem with the limitation imposed by the “no graver harm” component – it is nothing less than a legal requirement of heroism. What the law demands in the circumstances of “graver harm” is that the defendant be willing to sacrifice his life and/or limb (or those of his loved ones) in order to comply with the law – to willingly die in order to avoid killing an innocent or committing some other “graver harm.” It is difficult to see this requirement as anything other than unrealistic and absurd; a law that only saints and martyrs could possibly obey.⁵⁷

Post-*Erdemović*, it seems fairly well established in contemporary jurisprudence (despite some contradictory past case law) that duress cannot excuse murder, as the harm caused by the defendant taking an innocent life

⁵⁴ Cassese and Gaeta (2013) at p.215; this standard of duress emerges from the case law prior to the ICC, but there is at this stage no indication that the ICC would seek to depart from it.

⁵⁵ *Ibid*

⁵⁶ J. Ohlin (2009) “Necessity and Duress,” in A. Cassese *The Oxford Companion to International Criminal Justice*, Oxford: Oxford University Press, at p.432. This passage is cited as setting out the key criterion in making out duress by Cassese and Gaeta (2013) at p.215

⁵⁷ Cassese and Gaeta (2013) at p.215

is considered *prima facie* graver than any harm that the defendant seeks to avoid.⁵⁸ In *Erdemović*, the credible threat against the life of the defendant (which he interpreted as also a threat against the lives of his wife and child) was insufficient for the majority judges to afford the defendant duress as an excuse to liability, given that the crime that he committed was killing about 70 people. But as Judge Cassese argued forcefully in his dissenting opinion, this was not an inevitable conclusion based on past case law. He examined a number of post-WWII cases which could stand as authority for the proposition that duress can excuse even taking a life, including *Jepsen*,⁵⁹ where the Judge Advocate explicitly acknowledged that duress would be a defence if its conditions were met (though on the facts, it was not proven),⁶⁰ and *Einsatzgruppen*,⁶¹ where the Military Tribunal was explicit in its approval of duress as a defence to unlawful killing:

Let it be said at once that there is no law which requires that an innocent man must forfeit his life or suffer serious harm in order to avoid committing a crime which he condemns...No court will punish a man who, with a loaded pistol at his head, is compelled to pull a lethal lever.⁶²

However, as Judge Cassese acknowledged at para 43 of his opinion, whilst the cases he cited allowed for duress in principle, in only very few was it ever upheld on the facts in relation to unlawful killing. One of the relevant considerations in those cases where duress succeeded seems to have been whether the crime would have been committed anyway – so in a case where the killing was carried out by a group of people (such as an execution squad), the defendant’s participation or lack thereof would have made no difference to the outcome. In these circumstances, Judge Cassese asserts that the law cannot demand of a defendant that he “forfeit his life for no

⁵⁸ See the discussion in Cassese and Gaeta (2013) at pp.217-219.

⁵⁹ *Trial of Gustav Alfred Jepsen and Others*, Proceedings of a War Crimes Trial held at Lüneburg, Germany (13-23 Aug. 1946), Judgement of 24 Aug. 1946

⁶⁰ Separate and Dissenting Opinion of Judge Cassese, *Erdemović* at para 23; according to Judge Cassese, there was nothing in later authorities that would overrule the principle set out in *Jepsen*, contrary to what the Prosecutor and the majority claimed – see his discussion at paras 24-39.

⁶¹ *Trial of Otto Ohlendorf et al.*, (“*Einsatzgruppen*” case), in *Trials of War Criminals before the Nuernberg Military Tribunals under Control Council Law No. 10*, (U.S. Govt Printing Office, Washington D.C., 1950)

⁶² *Ibid*, p.480

benefit to anyone and no effect whatsoever apart from setting a heroic example for mankind.”⁶³ (emphasis in the original) Whilst this sounds deceptively sensible, it would mean that martyrdom would still be required of those who are unlucky enough to be coerced into killing in the absence of someone else to shift the blame on – hardly a satisfactory state of affairs.

3.3.3 Superior orders as an aspect of duress

A clear boundary is visible in contemporary law between the two defences, arising from the components included in the formulation of each. An invocation of superior orders requires “ignorance of illegality” of the conduct ordered, by the subordinate charged with carrying it out; whilst to prove duress, one must show that one’s acts were caused by a threat of either imminent death or serious bodily harm – an “irresistible exterior force that imposes a mental compulsion,” impairing the defendant’s free will.⁶⁴

However, this ostensibly clear separation can appear illusory, the respective legal formulations masking the fact that both defences concern in reality the same thing – the mental pressure that robs the defendant of any meaningful choice to do anything other than comply. The acknowledgement of that underlying reality can be noted in some of the post-WWII trials, where the tribunals considering the excuse of superior orders “focused on the existence of moral choice, at times effectively treating this plea as a military-specific doctrine of duress.”⁶⁵ One of the strongest expressions of this was the judgment of the Italian Court of Cassation in *Caroelli*, where the Italian Court of Cassation acquitted three defendants (Mr. Caroelli and two of his aides) who followed the order of their superior to execute ten partisans by way of unlawful reprisals. Despite there being no evidence of any threats being issued to make them comply, the court found that the defendants “lacked freedom of will, in the conduct ordered by their superior.” The judges accepted that Mr Caroelli initially

⁶³ Separate and Dissenting Opinion of Judge Cassese, *Erdemović*, paras 43-44

⁶⁴ G-J. A. Kooops (2008) *Defences in Contemporary International Criminal Law*, 2nd Ed, Leiden: Martinus Nijhoff Publishers, pp.43-44.

⁶⁵ C. K. Penny (2010) “Obeying Restraints: Applying the Plea of Superior Orders to Military Defendants before the International Criminal Court,” *Canadian Yearbook of International Law*, Vol. 48, pp.3-38 at p.15

tried to oppose the order and made his objections to his superior; and that when he left the superior's office, he had "a cadaverous appearance" and "could hardly stand on his feet," thus showing that the order put Caroelli in a state of "psychic confusion that was also accompanied by clear physical manifestations." This state was then transmitted to his aides. The court held that:

...when the manifestation of will contrary to the criminal action ordered by the superior is such as to cause clear physical troubles and a psychic confusion that nullifies the subordinate's freedom of decision, clouding a clear vision of hierarchical relations, evidently there does not exist that integrity of awareness and will required for making up a generic criminal intent, and even more the specific criminal intent necessary for the crime at issue.⁶⁶

In *Erdemović*, the line between the two defences was again blurred by the fact that the imposition of the superior order was combined by threats against the defendant's life. In its sentencing judgment, the first Trial Chamber found it difficult to disentangle the effects of the two, at one point appearing to bring into being a hybrid defence combining elements of both superior orders and duress.⁶⁷ At paragraph 19, the court appears to acknowledge that an order could have the effect of removing one's choice in certain circumstances, essentially amounting to duress:

...while the complete defence based on moral duress and/or a state of necessity stemming from superior orders is not ruled

⁶⁶ *Caroelli and others*, Italian Court of Cassation 10 May 1947, unreported at p2 cited in Cassese and Gaeta (2013) at p.235. Cassese argues that this was not in reality an instance of superior orders, but rather mental disorder that vitiated the defendants' liability. However, mental disorder was neither raised nor found, and would have been difficult on these facts to assert, given that it generally requires proof of a state that "destroys a person's capacity to appreciate the unlawfulness or nature of his or her conduct, or capacity to control his or her conduct to conform to the requirements of the law" (Art.31(1)(a) of the Rome Statute). See for instance, *Stenger and Crusius*, Leipzig Supreme Court, 6 July 1921, *Verhandlungen*, 2563-72 (cited also in Cassese and Gaeta (2013) at p.225), where a mental state amounting to "extreme agitation and psychological suffering" was apparently not sufficient to preclude the defendant's "free determination of will" (whatever that might be). It was only from the time that the defendant, later that same day, started running around, screaming, "desperately uttering calls, and leaving the overall impression of a maniac" that the court was willing to accept that his mental condition was sufficiently grave to preclude liability. By those standards, Mr Caroelli did not suffer from mental disorder; indeed, it's difficult to see how anyone not behaving like an inmate of a secure mental institution ever would.

⁶⁷ Sentencing Judgment, *Erdemović* (IT-96-22-T), Trial Chamber I, 29 November 1996, paras 15-19

out absolutely, its conditions of application are particularly strict. They must be sought not only in the very existence of a superior order - which must first be proven - but also and especially in the circumstances characterising how the order was given and how it was received. In this case-by-case approach - the one adopted by these post-war tribunals - when it assesses the objective and subjective elements characterising duress or the state of necessity, it is incumbent on the Trial Chamber to examine whether the accused in his situation did not have the duty to disobey, whether he had the moral choice to do so or to try to do so. (*my emphasis*)

Whilst the Appeal Chamber subsequently confirmed that no such hybrid defence exists in law,⁶⁸ the initial confusion is useful in illuminating the real issue at the heart of both defences – the absence of a realistic choice due to mental compulsion. If that is right, and the ability to choose not to commit the crime is necessary for criminal liability, then whether the absence of that ability is brought about by the coercive effect of an order from one’s superior, by threats, or by some other means, should matter little; it is the effect of these components on the defendant’s ability to choose that is the only thing that ought to be considered. Seen in this light, the different components of the two defences – the lack of knowledge of illegality and the necessity of threats against one’s life – seem to merely deflect from the reality of the absence of choice that the court is actually looking for, by pretending either that what is needed is for the choice to not arise at all (the soldier did not know his action to be illegal, and thus it was right of him to give into the pressure to obey) or that only a threat against life and limb is sufficient to cause sufficient pressure to remove one’s ability to choose. The limits thus imposed not only impose an artificial separation between the two defences, but also exclude from their scope those defendants whose inability to make a meaningful choice was brought about by circumstances other than those within the narrowly prescribed requirements of duress.

⁶⁸ Joint Separate Opinion of Judge McDonald and Judge Vohrah, *Erdemović* (IT-96-22-A), 7 October 1997, paras 34-35

3.4 What is missing from the law?

As evident from the above, ICL retains a strong focus on individual criminal responsibility, and it is reluctant to excuse the physical perpetrator from liability. Resort to such defences as superior orders or duress is therefore strictly limited. The culpability of those higher in the chain of command, whilst clearly acknowledged by such doctrines as command responsibility and indirect perpetration, is limited to liability for crimes that the superior knew about (either actually or constructively) and either failed to prevent or failed to punish; or indirectly perpetrated himself through his subordinates. Whilst this last doctrine of indirect perpetration appears to at least acknowledge the causative effect that an authority can exert upon subordinates, it is limited to liability for crimes that the superior intends and plans to commit.

However, as the discussion below will attempt to demonstrate, with any group enterprise there is always a substantial danger that immoral and criminal conduct can arise not through any intentional, explicit design of particular individuals within the group, but rather simply as a result of the group's particular attitudes, beliefs and policies (often all the more powerful for being unspoken), entrenched in its structures and practices. While domestic criminal and civil law have come to recognize and seek to prevent this possibility to a certain extent – by imposing vicarious and strict liability duties upon organizations to ensure that they develop effective procedures and training as means of preventing undesirable behaviour amongst its employees – international criminal law is yet to address this problem in any meaningful way. The people liable for a crime are still, in ICL, only the people directly involved in the crime – only those who either physically perpetrated the crime, planned or intended it to happen, or known about it in some way and failed to prevent or punish it. This ignores those responsible for shaping the perpetrator's mindset and propensity to commit crime through training and socialization into the group, even when such indoctrination makes it all but inevitable not only that some of the group members will likely commit an atrocity, but that others in the group - instead of stopping the perpetrators - will either join them or seek to cover

up the crime (or both). Part of the reason for this appears to be that the jurisprudential understanding of human behavior and choice has barely altered since the 19th century, and has not kept up with discoveries that have been made in fields dedicated to the study of this subject – as will be further explored in the next Chapter.

4. Understanding human behavior

”Psychology and law share a basic preoccupation: understanding the nature of human thought and action.”⁶⁹ Much like psychology and other behavioral sciences, international criminal law (and criminal law in general) dedicates a great deal of attention to human behavior and its causes, seeking to understand and deter conduct that is considered undesirable by the community. That preoccupation is observed in the very structure of criminal law. The prohibited physical act (the *actus reus*) is not, in general, sufficient for the imposition of criminal liability, but must be accompanied by the requisite state of mind (the *mens rea*) – the thoughts, intentions and other mental states that the law deems to have led to or caused the criminal behavior in a way that makes it just to hold the perpetrator culpable. When judges pronounce on whether a defendant before them was in control of his actions at the time he killed his victims, or when legal scholars discuss the deterrent effect – or like thereof – of imposing a particular punishment, they are engaging in an examination of the mental states and attitudes that cause or prevent a particular behavior; in effect, they are engaging in the field of psychology. It is therefore important to examine what the legal understanding of that psychology is and examine its validity against the evidence of human behavior emerging from behavioral science.

4.1 The legal understanding of human behavior

It is generally acknowledged that overall the law appears to retain the underlying assumptions of mental/physical dualism (the physical body and brain being somehow separate from the mind), with the mind imagined as a unitary organ of thought that controls the body. The average jurist’s understanding of what the mind is and how it works was aptly described by one scholar through the metaphor of the homunculus:

⁶⁹ O. R. Goodenough (2006) “Can Cognitive Neuroscience Make Psychology a Foundational Discipline for the Study of Law?” in *Law and Psychology: Current Legal Issues Volume 9*, B. Brooks-Gordon and M. Freeman eds. Oxford: Oxford University Press at p.78

The image is of a little rational figure at the centre of our thinking, like a captain on the bridge of a ship, barking out orders to a sometimes emotional, befuddled, and inattentive crew.⁷⁰

The conscious, rational mind – the captain of the ship – commands the behavior that the body will engage in. Therefore, that behavior is brought about primarily by the rational choices of an individual, who is a reasoning agent capable of appraising his options and choosing logically the one he prefers. It is recognized that this rational capacity can be undermined by mental illness or some other severe impairment – a physical brain injury for instance, or extremes of duress – causing the captain to lose his grip on the helm, and making “moral choice” impossible. However, in the face of any “lesser” challenges – emotions, stress, or other pressures – the law expects the captain to retain control of his vessel, and proceeds from the assumption that the rational mind remains the ultimate arbiter of what the body will do.

That these general assumptions also persist in the minds of judges and legislators who shape international criminal law is evident in the case law that considers crimes committed under social and other pressures. A number of passages from judgments in such matters testify to the fact that judges generally consider soldiers to be primarily rational agents who, in the moment they choose to commit a crime, exercise a rational moral choice – the captain of their mental ship chooses to give in to the pressure that they are under, and could have realistically chosen not to. For instance, the court in *US v Hutto* (concerning the My Lai massacre in Vietnam) said that

[A] member of the United States Army is not and may not be considered, short of insanity, an automaton, but may be inferred to be a reasoning agent who is under a duty to exercise moral judgment in obeying the orders of a superior officer.⁷¹

In a similar vein, the US war crimes tribunal at Nuremberg described the a soldier as a “reasoning agent” who “does not respond, and is not expected to

⁷⁰ *Ibid.*

⁷¹ Green (1976) at p.135.

respond, like a piece of machinery.”⁷² The fact that the courts see the moment when the perpetrator commits the crime as a product of a rational conscious choice – in which the perpetrator rationally weighed his options, considering the possible consequences that flow from each alternative – also found a clear expression in the *Hostages case*, where the court said that

It is true that the foregoing rule compels a commander to make a choice between possible punishment by his lawless government for the disobedience of the illegal order of his superior officer, or that of lawful punishment for the crime under the law of nations. To choose the former in the hope that victory will cleanse the act of its criminal characteristics manifests only weakness of character and adds nothing to the defence.⁷³

The law further assumes that in the moment that the choice is made, the perpetrator has access to some objective moral knowledge he knows to be applicable to his current circumstances, and which is necessary to make a rational decision – such as, for instance, awareness that killing the particular enemy civilians he is asked to execute is both wrong and illegal. Whilst the courts do not expect detailed legal knowledge of the technicalities of international criminal law, in circumstances where morality is objectively straightforward, the judges seem to believe that the perpetrator’s own little Captain Rationality at the helm must have access to the knowledge that the contemplated action is wrong (or more wrong than the alternative) – and decides to proceed anyway in the face of that knowledge. For instance, the Judge Advocate in the *Peleus Case* expressed his disbelief that it was not clear to the perpetrators what the right course of action was in the following way:

If this were a case which involved the careful consideration of questions of international law as to whether or not the command to fire on the helpless survivors struggling in the water was lawful, you might well think it would not be fair to hold any of the subordinates accused in this case responsible for what they are alleged to have done; but is it not fairly obvious to you that if in fact the carrying out of Eck’s command involved the killing of these helpless survivors, it

⁷² *In re Ohlendorf (Einsatzgruppen Trial)* (1948), 4 Nuremberg Mil. Tribs. 470, quoted in Green (1976) at pp.312-313

⁷³ *In re List* (1948) 8 War Crimes Rep. 34, at 50-52, quoted in Green (1976) at p.309

was not a lawful command, and that it must have been obvious to the most rudimentary intelligence that it was not a lawful command...⁷⁴

The belief that it must be obvious to any sane person, no matter the circumstances, that certain actions they carry out – even under pressure - are by their very nature morally wrong and criminal was expressed even more clearly by the Israeli court in *Eichmann*, talking about when a perpetrator must realize that the order he was given was illegal:

The distinguishing mark of a “manifestly unlawful order” should fly like a black flag above the order given, as a warning saying “Prohibited.” Not formal unlawfulness, hidden or half-hidden, nor unlawfulness discernible only by the eyes of legal experts is important here, but a flagrant and manifest breach of the law, definite and unnecessary unlawfulness appearing on the face of the order itself, the clearly criminal character of the acts ordered to be done, unlawfulness piercing the eye not blind nor the heart stony and corrupt – that is the measure of “manifest unlawfulness” *required to release a soldier from the duty of obedience upon him and make him criminally responsible for his acts.*⁷⁵ (my emphasis)

Therefore for jurists, behavior is the result of a conscious choice made by a reasoning agent, who has at his disposal sufficient knowledge to make his choice – and the resulting behavior – rational. Whilst the law might acknowledge the pressure the defendant faces when making his choice (be it threats, orders, or the more subtle influence of peer pressure), its model of human behavior is predicated on the belief in the ability of our rational minds to tame and overcome such pressure – with those unable to do so being judged as weak or criminal. In this way, the law is able to assign primary blame for the crime committed to some essential defect in the person who committed it, such as a weakness of character (as the court the in *Hostages* case explained), enabling it to justify punishment of that person.

⁷⁴ Quoted in Green (1976) at p.285

⁷⁵ From *Chief Military Prosecutor v Melinki*, quoted in *Eichmann*, which in turn was quoted in Green (1976) at p.103

When the judges in *Erdemović* spoke of “moral choice,” they spoke from the perspective of this legal model of rationality. It is therefore likely that to them, Erdemović made a rational, and even logical, decision by thinking through the anticipated consequences of his options, and chose the one that offered the better outcome for him and his family. The pressure of the order and the subsequent threats were, for the law, only one of the factors that Erdemović had to consider at the time, and one that he had the moral and legal duty to overcome.

4.2 The reality of human behavior

If people are primarily reasoning agents, and the choices that they make are as rational as the legal model of human behavior would have us believe, then the behavior of the majority can be predicted on the basis of what rational options exist in a given situation. For instance, when politely asked to slowly torture someone for a small monetary compensation, a reasonably moral rational agent – knowing full well that torture is morally wrong and illegal, and he could get in trouble with the law – would refuse to do so, no matter how insistent the request. If the legal model of rationality were correct, only a small percentage of individuals – those incurably criminal, sadistic or insane - would agree to carry out the action requested.

In the early 1960s, the psychologists polled on the subject agreed – less than 1% of people in an average population would be sufficiently deviant to commit such an act.⁷⁶ Therefore, when in his famous experiment Stanley Milgram persuaded over 60% of perfectly ordinary people who volunteered to be his subjects to – as far as they knew – torture a person to death with electric shocks, the childish ease with which he accomplished this task forced a reevaluation of the alleged rationality of normal human behavior. His research, together with that of Philip Zimbardo and Solomon Asch, exposed the power that social influence – conformity, obedience and institutional pressure – has over everyone’s behavior.

⁷⁶ Gilovich et al (2016) at p.337

4.2.1 Conformity

In one of the best-known set of experiments in social psychology, Solomon Asch demonstrated that even where the correct and rational action seems objectively obvious, the majority of people will conform their behavior to that of the group they are in – no matter how stupid or irrational that behavior might be. His subjects were asked to perform a simple task: to choose which of the three lines on a page matched a target line. There was no ambiguity about which of the choices was correct – the other two lines were clearly too short or too long to be a match – and when the participants in the control group made these judgments on their own, they almost never made a mistake. The participants were asked to call out their answers one by one, in groups of eight.

However, in each group of eight there was only one real subject – the others were “confederates” of the experimenter, and made their selections in accordance with his instructions. Asch wanted to test what the real subject will do when their own perception of reality clashes with the choice made by the rest of the group – will the subject select the obviously right answer, or will he conform to the answer given by others?⁷⁷

In general, when the confederates started giving the same wrong answer, the real subjects conformed to the opinion of the group – three quarters of the subjects conformed at least once. Asch’s experiment, as well as follow up studies, confirmed that both group size and unanimity is important – to achieve maximum conformity, the number of individuals reporting the same incorrect answer must be at least four (lesser rates of conformity are observed with fewer group members). When the subject’s own private opinion is supported by another group member’s dissent (not necessarily by giving the correct answer, just one different from the rest of the group), or he is given a chance to answer anonymously, conformity drops to around 5%. The individual’s culture also has an effect – individuals from more inter-dependent cultures tend to conform more than those from cultures where individuality is more valued – as does the perceived

⁷⁷ Description of the experiment and its results from Gilovich et al (2016) at pp.311-312

expertise or status of group members giving an opinion (greater status equals greater conformity).⁷⁸

That people conform when judging lengths of lines on a piece of paper might not seem terribly alarming, but unfortunately the same effect has been observed in relation to choices that are far more important – including those potentially concerning life and death – particularly when it is inaction, rather than action, that is dictated by the group. In one study,⁷⁹ nine out of ten subjects would not leave a room that started filling up with smoke when two other people in the room (confederates of the experimenter) remained seated at their desks, calmly filing out a questionnaire – although when no others were present, three-quarters of the subjects did go out to report the smoke. The subjects who remained in the room clearly acknowledged the smoke – they would cough, rub their eyes, and even open a window – but they did not do what the obviously rational thing to do was: leave the room, in case the smoke was an indication of a fire in the building. If this had been a real emergency, they would likely all be dead.

Why did people conform? Not unexpectedly, some of Asch's subjects later reported that they knew what the right answer was, but that it was easier to go along with the rest of the group. However, a more worrying trend emerged with others, who reported a belief that the group was right and their own initial perception was in error, suggesting that the group's influence actually had the power to alter their perception of reality. A later experiment conducted using fMRI⁸⁰ and a similar paradigm provides some support to this conclusion, showing that in some of the subjects, there was no activity in those areas of the brain that are responsible for monitoring conflict when the individual yielded to the opinion of the group – activity one would expect to find if the individual was struggling to

⁷⁸ *Ibid*, pp.314-320

⁷⁹ B. Latané and J. M. Darley (1968), "Group Inhibition of Bystander Intervention in Emergencies," *Journal of Personality and Social Psychology*, 10(3), pp.215-221

⁸⁰ fMRI stands for "Functional magnetic resonance imaging," a technique that measures brain activity by detecting changes in blood oxygen levels in the brain (which arise in association with such activity), using a strong magnetic field. The technique is cogently explained in O. R. Goodenough and M. Tucker (2011), "Neuroscience Basics for Lawyers," *Mercer Law Review*, 62(3), pp.945-958

reconcile his own perception of reality with that of the group.⁸¹ Further, the people in the “smoke” experiment who remained in the room later denied that they paid any attention to the actions of others present in the room, and reported instead that they did not see the smoke as an indication of fire (even thought that was the most obvious explanation). They instead came up with an inventive array of non-dangerous alternatives like steam, fog or “truth gas.”⁸²

4.2.2 Obedience

Stanley Milgram’s obedience experiments took the understanding of the power social influence a step further.⁸³ Milgram discovered that most people will obey an order to hurt another person – provided that the order comes from an authority they recognize as legitimate - even when that order is not backed up by any sanction or threat. In his initial experiment, Milgram recruited a number of participants from an average American town, ostensibly for a study about learning. When each subject arrived at Yale University, where Milgram’s lab was located, he would joint another participant, and the two of them – after drawing lots to determine who was to be the “teacher” and who was to be the “learner” – were briefed about the experiment by a scientist in a grey coat. The scientist would explain that the study was about the effect of pain on the ability to learn, and that the job of the “teacher” will be to administer an electric shock every time the “learner” gave a wrong answer – starting from 15 volts, and increasing by 15-volt increments with every wrong answer until 450 volts. To give the “teacher” an idea of how much the shocks hurt, he administered a 45-volt shock to him; then took him to another room, where the “shock generator” machine was. The machine had switches labeled in 15-volt increments, with

⁸¹ P. Zimbardo (2007) *The Lucifer Effect: How Good People Turn Evil*, Random House, London at pp.263-265

⁸² Latané and Darley (1968), pp.215-221

⁸³ The following description of Milgram’s experiments is summarized from Gilovich et al (2016) at pp.9-11, and 333-337; S. Milgram and T. Blass eds.,(2010) *The Individual in a Social World: Essays and Experiments*, 3rd ed., Pinter & Martin, London at pp.120-131; and Zimbardo (2007) at pp.266-271. Videos of the experiment are available of YouTube, for instance at <https://www.youtube.com/watch?v=eTX42IVDwA4>, accessed on 12 May 2016.

descriptions such as “Intense Shock,” “Extremely Intense Shock,” “Danger, Severe Shock” and “XXX” towards the higher end of the scale.

The teacher could hear the learner, but could not see him. As the experiment progressed, with the teacher delivering increasing levels of shock, the learner increasingly started to protest. At 150 volts, he wanted the teacher to stop; with each increasing shock, he would protest, scream, demand to be released, complain of his heart trouble, and refuse to answer, until finally, he would fall ominously silent. In the face of such protests, the teacher naturally wanted to stop; but then the scientist would prompt the teacher to continue shocking the learner, using one of four phrases – “Please continue,” “The experiment requires that you continue,” “It is absolutely essential that you continue,” and “You have no other choice; you must go on.”⁸⁴

An objective observer would note that the teacher clearly had a choice – no threats were delivered and no great reward was in jeopardy if he didn’t successfully complete (other than the small payment he had already received for participation). The door was not locked, and theoretically, the teacher could have just got up and walked out. Despite this, 62.5% of subjects in the initial experiment continued pressing the levers, shocking the learner – whose suffering, protests, and then possible death they were listening to – with shocks up to 450 volts, with no more apparent coercion than a few polite words from the scientist. The rate of compliance rose even higher – to between 72.5% and 92.5% - if other “teachers” who complied with the scientist without demur or chided the subject for his reluctance were present.

Milgram’s experiments were replicated all over the world, with similar results.⁸⁵ Further studies, and real-life situations, demonstrated the widespread application of his research. For instance, in one experiment, between 54% (men) and 100% (women) of participants exhibited maximum compliance (450 volts) with instructions to torture a puppy with real electric

⁸⁴ In reality, nobody was actually getting shocked and the “teacher” was the only subject of each experiment - the “learner” and any additional “teachers” were confederates.

⁸⁵ Zimbardo (2007) p.275 for a summary of experiments conducted in the US, Europe, Africa and Asia between 1963 and 1985.

shocks;⁸⁶ and all but one of the nurses who were the unwitting subjects of another experiment complied with telephone instructions from a stranger (who identified himself as a doctor) to administer twice the maximum amount (indicated on the label) of medication that could be lethal to their patients.⁸⁷ In real life “strip-search scam” cases, managers and employees in sixty eight different fast-food restaurants across the US complied with instructions of a stranger, who identified himself as a police officer, to strip-search an attractive employee (who the caller claimed was suspected of theft), carrying out increasingly degrading and humiliating sexual acts upon their unwilling subordinate as the caller’s demands escalated.⁸⁸

These results cannot meaningfully be described as actions brought about as a result of any kind of rational choice – unless one is willing to accept that the vast majority of the human species (one’s own self almost certainly included) is either evil or exhibits the kind of criminal “weakness of character” that ICL judges so disdain. What these experiments demonstrate is that almost anyone can be easily persuaded to commit an atrocious act, no matter how much they might not want to, in circumstances where it makes no rational sense for them to comply. It was very obvious that the subjects of Milgram and others did not want to do as asked – Milgram’s participants would protest, point out that the learner was in pain, and plead with the scientist to let them stop. Some of the subjects became quite distressed, stuttering, sweating and bursting into nervous laughter. Others would attempt to defy the scientist – they would even get up from their chair, only to sink back in defeat and continue pressing the levers. And yet, despite the fact that there was – objectively speaking – nothing seemingly capable of coercing their compliance, ultimately very few could break free; as one commentator put it, they “tried to disobey, but they weren’t particularly good at it.”⁸⁹ The irresistible effect that can be exerted

⁸⁶ *Ibid*, pp.275-276. They could see and hear the puppy suffering in response to the shocks they administered, though the shocks were in reality lower than those labeled. The subjects were clearly upset by this, and some of them cried – and kept on shocking the puppy anyway.

⁸⁷ See Zimbardo (2007) at pp.272-276; he notes that the nurse who refused to comply should have gotten a “raise and a hero’s medal,” so rare is the ability to resist obedience.

⁸⁸ *Ibid* at pp.277-288

⁸⁹ Gilovich et al (2016) at p.338

by authority was vividly described by one of the observers of the Milgram experiments⁹⁰ in the following way:

I observed a mature and initially poised businessman enter the laboratory smiling and confident. Within twenty minutes he was reduced to a twitching, stuttering wreck, who was rapidly approaching a point of nervous collapse. He constantly pulled on his earlobe and twisted his hands. At one point he pushed his fist into his forehead and muttered: "Oh God, let's stop it." And yet he continued to respond to every word of the experimenter and obeyed to the end.⁹¹

4.2.3 Institutional influence

Whilst in the previously described experiments the pressure that caused the behavior could at least be identified with some confidence – the order of the authority, the unanimous decision of the group – Philip Zimbardo's Stanford Prison Experiment showed that no such express or distinctly identified source of pressure is needed; merely being asked to fulfill a role within an organization or institutional setting can be enough to cause truly bizarre (from the point of view of rationality) behavior to occur.

Zimbardo⁹² recruited 24 middle-class, educated young men who were undergraduates or summer students at Stanford University, and randomly assigned them half of them to be "guards" and half to be "prisoners." All of the participants were well aware that they were only playing a role. The guards were issued uniforms and reflective sunglasses; the prisoners had a chain locked around one ankle and wore simple tunics. The guards were given no instructions or commands that would require them to abuse the prisoners (and in fact, were prohibited from using any physical violence); they were only instructed to maintain law and order and prevent any escapes.⁹³ Yet, upon assuming their roles, most of them commenced verbally and physically humiliating the prisoners, quickly escalating to requiring the prisoners to wear bags over their heads, stripping

⁹⁰ *Ibid*, at p. 11: Observers were invited to view the proceedings from behind a one-way mirror, after other researcher in the scientific community expressed skepticism about whether the subjects actually believed they were shocking anyone.

⁹¹ *Ibid*; quoted from S. Milgram (1963), "Behavioral study of obedience," *Journal of Abnormal and Social Psychology* 67, pp.371-378 at p.377

⁹² The following description of the experiment is summarized from Gilovich et al (2016) at p.9; and Zimbardo's day by day description of events in Zimbardo (2007) at pp.30-194

⁹³ Zimbardo (2007) at p.31

them naked and making them engage in simulated sex acts; while the prisoners transformed into, as Zimbardo put it, “pathological victims.”⁹⁴ Those of the guards who did not engage in such behavior did nothing to stop those who did. The experiment had to be terminated on the sixth day (it was supposed to run for two weeks), due to the sadistic behavior of the guards and the distress of the prisoners.

These were not real guards or real prisoners – they were intelligent students at one of the best universities in the world, who probably never dreamed that they would be capable of the behavior they exhibited in the experiment. And yet, merely assuming a role was sufficient for most of them to conform to the expectations of that role, as if they could see nothing beyond it. Zimbardo himself was affected by his own role as the scientist in charge of the experiment and “superintendent” of the prison. Despite observing the young men he was in charge of abusing one another, it did not occur to him to interfere – his role was to observe and supervise, and so that was all that he did. His reactions were two-fold, in accordance with the dual roles he occupied. On the one hand, he “morphed into a Prison Authority Figure...the high-status, authoritarian, overbearing boss man,” the very qualities that he had opposed and detested all his life.⁹⁵ On the other hand, he observed the boys’ abuse of each other with the distance of a scientist – their behavior seemed notable to him only for providing interesting, even exciting, experimental results, rather than for its cruelty.⁹⁶ Compounding the problem was the number of “outsiders” who visited the prison – police officers, Zimbardo’s colleagues, a priest and a lawyer, and even some of the parents of the experimental subjects. They observed the suffering within it, and mounted no challenge against what went on – thus lending a sense of normalcy and even social approval to the situation. Without any effective challenge, the behavior of the subjects – and the experimenter – remained dictated by their roles in the situation that they found themselves in. As Zimbardo put it:

⁹⁴ Zimbardo (2007) at p.180

⁹⁵ *Ibid*

⁹⁶ See for instance, the description in Zimbardo (2007) at pp.168-171, narrating also the confrontation with a visiting psychologist that finally knocked him out of his complacency and ended the experiment.

It is evident that one does not appreciate the power of Situations to transform one's thinking, feeling, and action when caught in its grip. A person in the claws of the System just goes along, doing what emerges as the natural way to respond at that time in that place.⁹⁷

Because of the obvious ethical concerns, Zimbardo's experiment has never been properly replicated, and Zimbardo himself – possibly as a result of his very personal experience during the experiment – can be perceived as a little evangelical on the subject of the power of situations. However, behavior akin to that of the subjects of the Stanford Prison Experiment has been observed in similar real-life situations – most notably, in Abu Ghraib prison in Iraq, where the hooding and sexual humiliation of prisoners by the guards (such as stripping and hooding prisoners, as well as forcing them to perform simulated sexual acts) eerily resembled the actions of Zimbardo's subjects, some 30 years prior.⁹⁸

4.3 “*The traitor in your skull*”⁹⁹

The counter-intuitive nature of the above findings makes most people react in disbelief. They are unable to imagine themselves doing what the participants of Milgram or Zimbardo's experiments had done, and – unable to internalize the research findings - tend to overestimate their (and other people's) ability to resist. As a result, despite the clear evidence that the vast majority, given the right circumstances, will comply and conform (no matter how much they might not want to), some legal scholars cling tenaciously to the tiny minority of those able to resist as evidence that a rational choice, taking into account legal and moral norms, is possible. Implicit in their writings – and in the jurisprudential passages cited above – is the belief that if those who gave in only tried a little harder, if their wills were just a little stronger, they too would be able to join the exclusive club of “reasoning agents” able to defy orders or any other pressure. The problem is that there is no real evidence that the decisions of those who

⁹⁷ *Ibid*

⁹⁸ Gilovich et al (2016) at pp.6-7. See also the detailed analysis in Zimbardo (2007).

⁹⁹ Title of this section borrowed from K. Taylor (2004), *Brainwashing: The science of thought control*, Oxford: Oxford University Press, at p.103

resist are made on the basis of any rational, reasoning process either; and that, even if a small minority is able to think rationally, the fact of the matter is that the great majority of people seems unable to do so in the face of social (or institutional) pressure.

It is doubtful that any model – such as the legal one - that explains human behavior as primarily caused and controlled by a reasoning mind is capable of accounting meaningfully for the results of the above experiments. After all, the hundreds of subjects who took part in them retained their reasoning capacity as far as the law would recognize – they were normal, even intelligent, people. They were not insane, or under duress (within the legal definition thereof), or even coerced in any way that could aid them in mitigation if their actions had attracted the attention of the criminal law. They all can be assumed to have had the basic moral knowledge necessary to grasp that torturing someone with electric shocks, or forcing them to perform simulated sexual acts is obviously both wrong and illegal. The law therefore has no way to explain their behavior, other than by reference to some inherent defect within them – their criminal desires, weak will or the like. But given the percentages involved, does that mean that nearly everyone is a budding criminal, just waiting for his chance to do evil?

The alternative is to acknowledge that the legal model of human behavior, with its little Captain Rationality able to control the ship, is simply wrong, and take a look at how the “normal” decision-making process actually works (“normal” in the sense of unimpeded by the sort of things that the law has already recognized as impeding rational thought – insanity, a certain type of duress, etc.). As the body of research discussed below indicates, it is actually the mutinous, often invisible, crew that runs our mental ship; with the poor captain, undermined and subverted at every turn, getting his turn at the helm rarely, if at all.

4.3.1 The reality of decision-making

One of the first to point out just how little of our judgments and decision-making resemble the products of a reasoning process was an influential social psychologist, Robert Zajonc, who in 1980 presented a

then-controversial paper on the primacy of affect – feeling over thinking – in human decisions. In it, he argued that affective judgments are basic, inescapable, difficult to verbalize, and tend to be irrevocable. It is these reactions that form the backbone of our everyday decision making process:

It is generally believed that all decisions require some conscious or unconscious processing of pros and cons. Somehow we have come to believe, tautologically, to be sure, that if a decision has been made, then a cognitive process must have preceded it. Yet there is no evidence that this is indeed so. In fact, for most decisions, it is extremely difficult to demonstrate that there has actually been any prior cognitive process whatsoever...We sometimes delude ourselves that we proceed in a rational manner and weigh all the pros and cons of the various alternatives. But this is probably seldom the actual case...Most of the time, information collected about alternatives serves us less for making a decision than for justifying it afterwards.¹⁰⁰

Zajonc's insights have increasingly been vindicated. Whilst several models have been proposed concerning how judgments and decisions are made, heuristics, cognitive biases, and affect are acknowledged to play a big role in all of them.¹⁰¹ Broadly accepted, and supported by much of the experimental data, is the idea that human decisions are the product of the interaction of two types of processes, which can be described as automatic and controlled.¹⁰² The automatic processes are those that are fast, intuitive, affective, and employ heuristics to reach an outcome; as such processes are unconscious, we are generally unaware that they exist, and we have no control over the outcomes they achieve. One example of an automatic process is stereotyping, which produces a quick judgment or reaction about how to treat a person we encounter based on a classification of that person into a category, and our existing attitude towards that category. The controlled processes are those that are conscious, articulated, effortful and slow – these are the processes that our conscious self identifies with, and

¹⁰⁰ R. B. Zajonc (1980). "Feeling and Thinking: Preferences Need No Inference." *American Psychologist*, Vol. 35, No.2, 151-175 at p.155

¹⁰¹ See discussion in Eysenck and Keane (2015) in Chapter 13: Judgment and Decision Making and pp.655-681 (concerning affect and cognition).

¹⁰² See Eysenck and Keane (2015) at pp.561-564 (Kahneman) and 606-611 (Evans), as well as pp.664-666 (Greene, in relation to moral judgment).

which allow us to follow rules, employ logic and exert some self-control over our emotions and instinctive reactions. An example of a controlled process is multiplying large numbers in your head, or filling out your tax return – mental actions which require time, attention, and energy.

For the sake of simplicity, Daniel Kahneman – one of the leading experts in the psychology of how we think – invites us to think about the automatic processes together as System 1 and the controlled processes as System 2.¹⁰³ For most day-to-day decisions, System 1 has a perfectly adequate, fast and automatic answer, arising either from our innate capabilities, or acquired through prolonged practice (i.e. the answer to $2+2=?$ does not require effort, having been rehearsed extensively in primary school; neither does driving a car on an empty road after many hours of practice).¹⁰⁴ Some of the mental actions of System 1 are completely involuntary – a person cannot prevent himself from “knowing that $2+2=4$ or from thinking of Paris when the capital of France is mentioned.” The others, while still arising or proceeding involuntarily, can be controlled through the exercise of System 2 (i.e. chewing – usually on automatic, but we can stop it when our attention is directed to that action).¹⁰⁵

Monitoring and control of System 1 reactions is one of the functions of System 2. As Kahneman describes it, System 2 is the one “that can follow rules, compare objects, and make deliberate choices between options. The automatic System 1 does not have these capabilities.”¹⁰⁶ However, the exercise of that control is undermined by several problems. The first is simply the problem of energy. Mental effort draws from the same limited pool of energy that fuels the body as a whole, and the exercise of System 2 requires that a lot of it be diverted to its operation. Thus, for instance, people attempting to solve a non-automatic math equation (ie. 17×24) will find it difficult to walk at the same time, and may attempt to shut off any input that

¹⁰³ There are no actual “systems” as such; the systems referred to are merely rough groupings of different cognitive processes, made to simplify the overall explanation of how thinking works: see D. Kahneman (2011), *Thinking Fast and Slow*, London: Penguin Books, at pp.28-30

¹⁰⁴ *Ibid.*, at pp.21-22.

¹⁰⁵ *Ibid.*, at p.21-22.

¹⁰⁶ *Ibid.*, at p.36

requires energy to process – they might close their eyes, and stop paying attention to what they hear or feel.¹⁰⁷ The effort required becomes even harder in cases where System 1 has already given an answer, which then must be overridden, as “the conclusion comes first, and the arguments follow.”¹⁰⁸ When a person’s energy is already low, there might not be enough to feed the voracious energy requirements of System 2. An illustration of this is supplied by a study of eight parole judges in Israel, analyzing 1,112 decisions made over of 50 days. During the course of a day, each judge would consider a number of applications for parole, with an average of just six minutes to make his decision. The study found that after each food break the judges had, the proportion of paroles they granted spiked to 65%, and that proportion steadily declined in the two hours until their next meal to about 0% just before the meal – when the tired and hungry judges, lacking the energy needed for the mental effort to consider the merits of each application, would fall back on the default position of denying the request for parole.¹⁰⁹

The second problem is we have no conscious awareness, and thus no insight, into how the majority of our decisions are actually made. Thus, System 2 might not even spot the fact that a decision requires correction – such as, for instance, in cases where System 1 provides a ready answer to a slightly different question than that which was asked, without System 2 noticing the substitution. This occurs in many instances where a decision requires undertaking an examination of the merits of different options, and System 1 will still supply an answer even though it is unable to actually undertake the hard work of doing the comparison needed. For instance, faced with a question of whether to invest in a particular stock – say, Ford – the question asked might be “should I invest in Ford stock” but the quick

¹⁰⁷ *Ibid*, at p.20-23

¹⁰⁸ *Ibid*, at p.45

¹⁰⁹ *Ibid*, at p.43-44; the authors of the study apparently checked for alternative explanations, but this seems to be the best possible one for the data. See: S. Danziger, J. Levav and L. Avnaim-Pesso (2011), “Extraneous factors in judicial decisions,” *Proceedings of the National Academy of Sciences of the United States*, Vol. 108, No. 17, pp.6889-6892. This study fits within, and is consistent with other studies on mental fatigue conducted by Roy Baumeister, summarized in Kahneman (2011) at pp.41-44

answer that comes back will be the answer to a different question – “do I like Ford cars?”¹¹⁰

The flipside is that in people with expertise relevant to the question asked, System 1 is able to supply an often- accurate answer without having to go through any laborious System 2 process of consciously accessing needed knowledge and working it out. Expert physicians able to diagnose a patient at first glance, chess-masters who are able to foresee how a game will end just by glancing at the board, fire-fighters who get out of a burning building without realizing why, just before the floor they had been standing on collapses – these are all “miracles” of intuition that are nothing more than a feature of System 1, which, after hundreds or thousands of hours of training, can access relevant information and trigger action before the conscious mind even has the chance to notice the cue in the environment that prompted System 1 to act.¹¹¹

This failure of conscious insight into our actions is also why people are often unable to explain the things that they did, and instead come up with plausible sounding rationalizations for their actions when asked to explain themselves. The people in the Latané and Darley experiment (the “smoke” experiment) rationalized their failure to leave a potentially life-threatening situation by explaining to themselves (and the experimenter) that they did not believe the smoke to be harmful; Milgram’s subjects came to blame the learners for their “stupidity” in failing to answer questions correctly to explain their own failure to resist;¹¹² and Zimbardo’s subjects came up with a variety of rationalizations for their actions – for instance, one of the most sadistic of the “guards” explained in the debriefing that he was running his own experiment, trying to test how much people would take before rebelling.¹¹³

Finally, the third problem is that our thinking is prone to a variety of systematic errors in how we seek out information, process it, and come to

¹¹⁰ *Ibid* at p.12

¹¹¹ *Ibid* at p.11

¹¹² Milgram and Blass (2010) at pp.120-131

¹¹³ Zimbardo (2007) at pp.192-194

conclusion.¹¹⁴ For instance, confirmation bias ensures that when evaluating a proposition (i.e. “happy people live longer”), “people more readily, reliably, and vigorously seek out evidence that would support the proposition rather than information that would contradict the proposition.”¹¹⁵ If the proposition is something that a person already believes in strongly – i.e. “God exists” or “abortion is always wrong” – the information that supports that belief will be sought out and readily accepted, whilst any disconfirming information will be criticized, and often dismissed or re-interpreted to fit in with one’s existing views. One MRI study even showed that when faced with such disconfirming information, the part of the brain that usually processes information virtually shuts down.¹¹⁶

Our reasoning process is vulnerable to all manner of such cognitive errors¹¹⁷, one of which is particularly relevant here: cognitive dissonance. Cognitive dissonance is the driving need to resolve two incompatible ideas or facts in a way that is consistent with our own sense of self – a positive version of who we are. For instance, someone who smokes (fact: “I smoke”) knowing that smoking is seriously detrimental to his health (fact: “Smoking will kill me”) will experience the dissonance between the two facts. The solution is obvious – “I’m an idiot who should stop smoking” – but this is not what usually happens, particularly when someone is already emotionally invested in the idea or activity in question (“smoking makes me feel good”). Instead, the person will attempt to resolve the dissonance by convincing himself that smoking is not really all that bad, that he doesn’t really smoke all that much, that the stress-relief he gets from smoking compensates for its detrimental effects on his health, etc. The interesting thing about cognitive dissonance is that a person in its grip may often try very hard to “proselytize” – convince other people that his way of thinking

¹¹⁴ It is unclear whether these errors affect both System 1 and 2. Kahneman seems to think that it is System 1 that is affected, with System 2 simply unaware of and unable to spot the errors – once it is pointed out, System 2 can account for it and correct the answer accordingly. Others disagree, and argue both automatic and controlled processes are affected. See discussion in Eysenck and Keane (2015) at pp.563-564

¹¹⁵ Gilovich et al (2016) at p.124

¹¹⁶ C. Tavis and E. Aronson (2007), *Mistakes were made (but not by me): Why we justify foolish beliefs, bad decisions and hurtful acts*, London: Pinter and Martin, at p.19

¹¹⁷ Discussing all of them is beyond the scope of this work; for an accessible guide to human irrationality, Tavis and Aronson (2007).

is correct, and that they too should accept it (“you should try smoking, it’s very relaxing”) – in order to gain social confirmation or approval for his habit (“everybody else around me is smoking, so it cannot be all that bad”).¹¹⁸

This presents an obvious problem when it comes to combatants committing atrocities – if a person who is emotionally invested in his own image of himself as a decent individual (as most of us are) tortures or murders someone, a dissonance will arise between his self-image and the atrocity he committed. Very few of us are able to resolve that dissonance by admitting that we committed a terrible act and may therefore not be as much of a decent person as we thought we were. Instead, the most popular way for people to resolve this kind of dissonance is to blame the victim – perversely, the more innocent the victim, the greater the dissonance, and the harder we must work to convince ourselves that they really deserved what they got, demonizing and dehumanizing the victim as much as we can.¹¹⁹ Worse yet, if this person who resolves their dissonance in this way is still in the midst of an ongoing conflict, he will now have a powerful incentive not only to commit further atrocities, but also to do everything possible to convince or force others to do the same. The same kind of dissonance will be at work in group members who are strongly committed to the positive image of their own group (because part of their sense of self is their membership of that group) – they will fight to exonerate group members who have committed

¹¹⁸ The psychologists who discovered cognitive dissonance infiltrated and studied a small doomsday cult, whose members believed that the world was coming to an end on a specific date. The scientists wanted to know what would happen to the members’ faith when it is conclusively proven to be false – when the given date comes and goes without the promised apocalyptic flood or the arrival of the benevolent aliens who were supposed to rescue the faithful from it. It turned out that, although some of the less committed members quit, the faith of the others became ever more fervent once they had successfully resolved the dissonance generated by the failure of the apocalypse to materialize (by convincing themselves the Earth was saved by the benevolent aliens, who were impressed by the cultists’ devotion). They also began to energetically proselytize in an active effort to recruit new members – something they were not interested in prior to their faith being disconfirmed. For the full account of the study, see L. Festinger et al (2008) *When Prophecy Fails*, London: Pinter & Martin

¹¹⁹ This is why Milgram’s subjects blamed the learners and called them stupid – they needed to resolve the dissonance between their own self-image and the fact they were willing to torture a perfectly nice, ordinary stranger who did nothing to them. This is not something that people consciously choose to do, merely an instinctive way we protect ourselves from injury to our version of “self.”

atrocities against the “enemy” or “stranger” group, and to convince whoever they can that the treatment meted out was justified or at least excusable. For example, polls indicated that 80% of the American public disapproved of putting Lieutenant Calley on trial for his part in the massacre of the village of My Lai in Vietnam, and 100,000 letters and telegrams were sent to President Nixon demanding that he be freed following his conviction; most common reason given for this support was the perceived unfairness of putting on trial a man who was only “doing his duty.”¹²⁰

4.3.2 “Moral choice”

The reason that moral choices and judgments feel so “right” and come to us so quickly is that they are primarily the result of the operation of System 1.¹²¹ No-one has to explain to most people that harming others is wrong – even small children who are yet to receive any kind of moral instruction know this to be true.¹²² At this point, there is quite a lot of research to support the idea that moral judgment is primarily instinctual and affective, with the relevant instincts arising from our genetic make-up and being shaped by our environment.¹²³ Two powerful instinctual responses are particularly relevant here: empathy (altruism) and in-group bias (ethnocentrism).

Empathy is what makes us care about other people; it’s the instinct that allows us to feel their pain, makes us want to help them, and stops us from doing harm. We are repelled not only from actions we can see causing

¹²⁰ Gonzalez (2004) at p.134

¹²¹ System 2 can also make moral judgments, and a minority of people have shown a propensity towards restraining their moral instincts and thinking through moral dilemmas presented to them in experiments, given sufficient time and energy. Unfortunately, the resulting (more utilitarian) moral judgments often resemble those made by psychopaths and patients with brain damage to a specific part of the pre-frontal cortex (VMPFC), whose moral System 1 instincts are impaired or lacking. See further: Eysenck and Keane (2015) at pp.664-667

¹²² For instance, 6-10 month old pre-verbal infants have been shown to make value judgments, preferring those they observe helping others over those whose actions are neutral or hindering – see J. K. Hamlin, K. Wynn, P. Bloom (2007) “Social evaluation by preverbal infants.” *Nature* 450, pp.557-559

¹²³ For a review of the studies showing that morality has an organic basis, see for instance L. Young and M. Koenigs (2007), “Investigating emotion in moral cognition: a review of evidence from functional neuroimaging and neuropsychology.” *British Medical Bulletin*, Vol.84, pp.69-79

harm to our victim (thus receiving instant feedback of victim distress), but even actions that are reliably associated with visible victim distress. For instance, in a recent experiment, simulating a variety of violent actions (ie. drawing a line across a victim's throat, pretend-smashing their leg with a hammer, or smacking a realistic baby doll against the table) was sufficient to stimulate an aversive physical response in the subjects – their peripheral blood vessels constricted giving them “cold feet” – although they had no problems performing a similar, non-harmful action such as hammering a nail into a block of wood. The subjects performed the “harmful” actions reluctantly, and as perfunctorily as they could, with one person simply refusing to do some of them at all.¹²⁴

Empathy is unfortunately limited in several ways, including by in-group bias – a powerful instinct that directs our empathy towards our own group or tribe, and limits or even extinguishes (in certain circumstances) empathy towards those outside our own group: a strong preference towards “people like us.” The instinct is universal – there are no societies yet discovered where outsiders are treated the same as insiders – and in-born, with babies as young as six months already showing a preference for “their own” kind of people.¹²⁵ There is some evidence that the mechanism of this in-group bias, particularly in a group conflict, is hormonal; with the hormone oxytocin promoting greater empathy towards those identified as in-group members at the same time as inhibiting empathy and promoting aggression towards those identified as out-group members (anything associated with each group – rituals, flags, or any other tribal markers – provokes similar reactions, respectively).¹²⁶ The aggression can be amplified

¹²⁴ On the other hand, we have no aversive reaction to actions that are not reliably associated with victim distress – such as pressing a button. Thus, the same person who would be able to activate a nuclear device from a distance, killing millions, might flinch from smashing a fake baby's head against the table. See further: F. Cushman, W. B. Mendes et al. (2012) “Brief Report: Simulating Murder: The Aversion to Harmful Action.” *Emotion*, Vol 12, No 1, pp.2-7

¹²⁵ J. Greene (2013) *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them* (Kindle edition) London: Atlantic Books at Loc 839: Six month old babies prefer to look at speakers without foreign accents; ten month old babies prefer toys from native language speakers; five year old children prefer to play with children who lack foreign accents.

¹²⁶ *Ibid* at Loc 903

by our tendency to dehumanize our enemies,¹²⁷ attributing “negative, nonhuman characteristics” to members of the enemy group. The greater loyalty we feel towards our own cherished social group, the greater is our tendency to dehumanize and be violent towards those who do not belong to that group.¹²⁸

It would thus be fair to say that we experience very different moral feelings and reactions when we deal with or think about out-group members than we do when we deal with members of our own group. Disgust,¹²⁹ anger and aggression are commonly felt towards out-group members who belong to a group that our own tribe is in conflict with or has traditionally exploited, or those who have transgressed the norms of our own tribe – particularly when something occurs to reinforce the in-group bias, such as a recent death of one of “us” at the hands of one of “them.”¹³⁰ At the same time, the empathy we would normally experience towards another person is likely to be dampened, or even completely absent, in relation to those belonging to the relevant out-group, particularly after we become desensitized by watching or performing abusive acts towards out-group members; while the empathy towards our own group is likely to become heightened. In those circumstances, it is fairly pointless to rely on some

¹²⁷ Or perhaps the feeling of aggression towards complete strangers produces dissonance, which is then resolved by dehumanizing the targets of that aggression, making the aggression justified.

¹²⁸ Summary of dehumanization from Gilovich et al (2016) at pp.501-502. Other strong inhibitors of empathy are: being in a position of power, de-individualization (i.e. being part of a crowd, particularly when covering one’s face or eyes), and desensitization (for instance, a person who might feel initial empathy towards their first victim will no longer feel any empathy towards subsequent victims, once killing becomes a matter of routine).

¹²⁹ Disgust is particularly dangerous when it becomes associated with a group of people, as it strongly motivates us towards elimination of whatever we find disgusting.

Dehumanizing language that invokes disgust in relation to a particular group – i.e. calling people “rats,” or “cockroaches” or “diseased” – primes the listeners to start associating members of the group with the feeling of disgust, paving the way for future cruelty towards members to that group. For details on the role disgust plays in atrocities, see K. Taylor (2009), *Cruelty: Human Evil and the Human Brain*, Oxford: Oxford University Press.

¹³⁰ Try, for instance, to examine your feelings towards people who have committed war crimes or those who abused children; most of us will be unable to feel any empathy towards them. If such people are abused, we might in the abstract acknowledge that abuse of any human being is wrong, but hardly anyone would feel particularly outraged about that abuse being applied to these particular human beings – many of us might even experience a kind of vicious satisfaction that they are getting a taste of their own medicine, so to speak. This is because, even though we might know they are human, they no longer feel human – they are excluded from the category of “Us” and placed into a particularly disparaged category of “Them.”

kind of common objective morality as a guide to what is or isn't grossly immoral or "manifestly illegal." The "knowledge" that comprises such sentiments is for the most part instinctual; and when the relevant instincts are extinguished towards a particular out-group, we may no longer "know" that slaughtering members of that out-group is wrong – it is difficult, if not impossible, to perceive as human those towards whom we feel no empathy. Researchers who have examined the psychology behind mass killings even go as far as to say that, as far as the subjective view of the perpetrators is concerned, they likely feel that the atrocity they are committing is morally right and just.¹³¹

That does not mean that everyone experiences these effects consistently or in the same way. Some people are naturally less aggressive¹³² and more empathetic than others; they may additionally have had their empathetic instincts reinforced by their upbringing and/or have had a lot of practice in caring for other people; or they may be naturally less tribal or have had experiences that make them less inclined towards ethnocentrism (such as extensive contact with many other cultures); or some combination of all of the above. For example, people who rescued Jews during the Holocaust, risking their own lives, did not tend to have any higher regards for Jews than those who did not provide assistance. Instead, they tended to have a "history of helping others" and reported a greater emphasis being placed on altruism and compassion in their homes when they were growing up.

Rescuers reported that their parents and grandparents frequently told stories from their culture in which altruism was a theme. Altruism was a central theme in the books the family read and the teachings they discussed. In their dinnertime conversations about the events of the day, they discussed things through the lens of altruism and concern for other people. Altruism was explicitly invoked as an important ethical principle.¹³³

¹³¹ Gilovich et al (2016) at p.514

¹³² For a brief explanation of the genetic and social factors that predispose towards aggression, see Gilovich et al (2016) at p.489.

¹³³ Gilovich et al (2016) at pp.332 and 529-530; see also S. Oliner and P. Oliner (1988), *The altruistic personality*, New York: Free Press, based on interviews with 100 rescuers from WWII.

Even in those who are not particularly altruistic, empathy can be triggered by particular circumstances even to the point of re-humanizing a member of the enemy group. For instance, soldiers have reported instances of forming an emotional connection with an enemy they had killed after finding family photographs amongst his effects – they become haunted by the faces in those photographs¹³⁴. However, what seems clear is that we have little control over when and towards whom such feelings may arise, and whether they will be strong enough to prevail over any competing instincts and direct our behavior in any given moment.

4.3.3 “Reasoning agents”

Soldiers – or any other combatants – are particularly likely to develop a strong in-group bias and a correspondingly weak sense of empathy towards anyone outside the in-group. This is due to a number of factors, starting with the fact that they are embedded in a closed society with its own rules and distinctive values – as the US Supreme Court acknowledged in *Parker v Levy* in 1974, it has long been recognized “that the military is, by necessity, a specialized society separate from civilian society.”¹³⁵ That separation, together with the training that a new recruit undergoes, as well as the perceived common goal and purpose amongst the members of the society (to defend the nation against a common enemy), serves to strongly reinforce group cohesiveness, a soldier’s identification with perceived group values, and his loyalty to and empathy with the members of his group. This can have a positive effect within the group – race and ethnic integration in the U.S. military are far more successful than in the country as a whole.¹³⁶ However, the correlate effects are increased in-group bias and thus weaker empathy towards the out-group; further, increased group cohesion positively correlates with a greater ability of the

¹³⁴ Gonzalez (2004) at p.138

¹³⁵ *Parker v Levy*, 417 U.S. 733 (1974) at 743

¹³⁶ Gilovich et al (2016) at p.408

group to ensure conformity with the group and its norms through social pressure.¹³⁷

Further, the training that soldiers receive is specifically designed to overcome their natural empathetic responses to actions associated with victim distress. It is not natural for anyone to kill complete strangers – but that is precisely what is part of a combatant’s job description. In past wars, this would present a problem when troops who were not effectively trained to kill would be reluctant to engage the enemy – one commander in WWII complained that his troops would only fire at the enemy if someone stood over them and yelled at them to do it; and apparently only 15-20% of riflemen in WWII fired their rifles at exposed enemy soldiers. This is why modern training (which the army employs social scientists to help design) consists of repeated simulations in mock battle conditions – “battle proofing” – to overcome resistance to killing through desensitization, get soldiers to react instinctually to certain stimuli, and expose them to the stress and dangers of combat. Such repeated exposure to battle scenarios eventually results in the trained reactions being carried out without thinking in response to the stimuli associated with those reactions.¹³⁸ No comparable behavioral training appears to be available to indoctrinate soldiers to comply with the laws of war or resist social influence – even the ICRC training in humanitarian law relies primarily on class-based education, and only recommends incorporating “small elements of law of war training” as part of normal field instructions or exercises.¹³⁹

In addition, the training and socialization that a soldier receives encourages blind obedience, in line with the social research that indicates that cohesive groups tend to react poorly to deviance of any kind from one of its members (“deviance” meaning anything other than blind conformity and obedience). Military officers in particular tend to be rated more

¹³⁷ See summary of research on compliance in the military in Gonzalez (2004) at pp.102-103 and 112-114.

¹³⁸ *Ibid*, pp.114-115 and 119-120

¹³⁹ D. L. Roberts (1997) “Training the armed forces to respect international humanitarian law: The perspective of the ICRC Delegate to the Armed and Security Forces of South Asia,” *International Review of the Red Cross*, No.319, found at <https://www.icrc.org/eng/resources/documents/misc/57jnrk.htm>, accessed on 26 May 2016

favorably if they follow a “no questions asked” approach, rather than engage in any critical thinking.¹⁴⁰ On the battlefield, any deviation from the group consensus can be met with violence or death – in one account mentioned by Gonzalez, a soldier who refused to participate in an execution of hostages was convicted of treason on the spot by the officer in charge, placed with the hostages, and shot by his fellow soldiers.¹⁴¹

Finally, the conditions that combatants endure in the midst of a group conflict can cause a significant deterioration in cognitive functioning. Such conditions include stress and anxiety, emotional and physical exhaustion caused by the repeated activation of the sympathetic nervous system (the “fight or flight” response), regular sleep deprivation (which, if it continues long enough, will cause hallucinations), as well as the effects of “malevolent environment” (bad food, poor living conditions, uncomfortable climate) which has been found to be a stressor for Post-Traumatic Stress Disorder even in absence of exposure to actual violence in combat.¹⁴² The resultant deterioration in function has been found to include tunnel vision, auditory exclusion, and the inability to think clearly.¹⁴³ Such cognitive deterioration will make it even more likely that a soldier will operate almost exclusively on the basis of his instinctual responses – which have been, through his training and socialization, heavily weighted towards an in-group bias and against any empathy towards those outside of that group, particularly if they are “the enemy.” It is therefore probably fair to conclude that – given the above factors – the only behavior that can reliably be expected by combatants in group conflicts is instinctual behavior, which falls far short of the behavior of a “reasoning agent” that the law expects.

¹⁴⁰ Gonzalez (2004) p.107

¹⁴¹ *Ibid*, p.108

¹⁴² *Ibid*, pp.123 and 141-149.

¹⁴³ *Ibid*, pp.141-142

5. Law as a deterrent

Deterrence – along with retribution for the crime, incapacitation of the offender, and rehabilitation – is generally presumed to be one of the chief goals of criminal law, both domestic and international. Whether in reality any deterrent effect of the law can be observed (and if so, to what extent and in what circumstances) is a matter of some debate, which is beyond the scope of this thesis. The discussion below will focus solely on the perceived deterrence effect of international criminal law on combatants committing their crimes in the context of a group conflict under the influence of social pressure.

5.1 The legal understanding of deterrence

In rejecting duress as an excuse to liability for murder – even in cases where the victim(s) would have died anyway had the accused opted for martyrdom instead of giving in to the pressure to kill them – the judges of the Appeal Chamber of the ICTY in the *Erdemović* case appear to have placed a great deal of emphasis on the role of law as a deterrent to crime.¹⁴⁴ Grounding their decision in a lengthy exposition of judicial precedent (primarily sourced from English common law) regarding the value of the law as an incentive to resist duress and conversely, the impunity that would result should that incentive be removed, Judges McDonald and Vohrah, in the Joint Separate Opinion, concluded as follows:

Our view is based upon a recognition that international humanitarian law should guide the conduct of combatants and their commanders. There must be legal limits as to the conduct of combatants and their commanders in armed conflict. In accordance with the spirit of international humanitarian law, we deny the availability of duress as a complete defence to combatants who have killed innocent persons. In so doing, we give notice in no uncertain terms that those who kill innocent persons will not be able to take advantage of duress as a defence and thus get away with

¹⁴⁴ See the Joint Separate Opinion of Judge McDonald and Judge Vohrah, *Erdemović* (IT-96-22-A), 7 October 1997, at paras 75-80.

impunity for their criminal acts in the taking of innocent lives.¹⁴⁵

Needless to say, this reliance on the deterrent effect of the law was not supported by any evidence that such an effect actually exist under the circumstances in question – namely, that knowledge of legal sanction has any effect in guiding the behavior of a combatant in the midst of a conflict, particularly one under pressure to commit the crime. Both in the opinions of Judges McDonald and Vohrah, as well as in the passages of case law they cited in support, it is simply assumed that knowledge of legal sanction (or lack thereof) forms part of a perpetrator’s deliberation over whether or not to commit the crime; and that his subsequent decision might be (or ought to be) affected by that knowledge. This is congruent with the assumptions made by earlier international criminal tribunals in the passages cited at 4.1 above: that the knowledge of illegality of criminal conduct can be presumed to have played a part in the perpetrator’s deliberations and his ”moral choice” to commit the crime. But here, the assumption goes a step further: the judges in *Erdemović* appear to be saying that the legal sanction in question will not only form part of future perpetrators’ rational decision-making processes, but that it might also have the effect of deterring at least some of them from committing their crimes.

5.2 The reality of deterrence

It should be apparent from the research summarized above that, contrary to the judicial understanding of these issues, human choice and human morality are generally not rational and rarely based on reason, at least outside schools of philosophy and law. Morality in particular is a complex set of emotions and instinctual reactions, shaped by our genetics and upbringing and triggered by circumstances. As with any System 1 mechanism, imposing System 2 control to consciously restrain our moral reactions requires time, energy and emotional distance from the problem at hand – all qualities that are in short supply during a group conflict. Further, conscious control requires awareness that our initial reactions may be

¹⁴⁵ *Ibid*, para 80

wrong, which in turn means developing a healthy suspicion of easy answers and group consensus, as well as perhaps a habit of introspection. This is something most people never learn, given that our moral and other System 1 instincts serve us quite well in everyday life, and soldiers in particular are not only never taught but positively discouraged from learning (given the military emphasis on group cohesion and instant obedience of superiors). Thus, relying on soldiers – or even their commanders, once any conflict begins – to behave like “reasoning agents” capable of exercising some kind of rational “moral choice” that takes into account considerations of international law seems to be more than a little unrealistic. In the midst of a conflict, their actions are far more likely to be governed by instinct, learned habits and reactions, and the social pressure exerted by their peers and superiors. It is therefore unlikely that the law will achieve deterrence as it currently aims to do so, by targeting the physical perpetrators of atrocities or only those superiors who knew about, and failed to prevent or punish, a specific crime.

5.3 Allocating liability to prevent atrocities

The behavioral research set out above indicates that it is unlikely that law will influence choices made once conflict begins, given that a set of abstract rules is unlikely to be able to compete with emotional, moral, and social instincts. It therefore needs to influence those decision makers who shape those instincts – those in charge of the military who decide what kind of training, socialization and discipline their soldiers are subject to.

At the moment, the usual type of training soldiers receive makes it more likely that they will commit atrocities – strong in-group bias in a closed society is combined with reinforcement of our social instincts to conform and comply, while at the same time behavioral training to kill makes it less likely the soldier will feel, or act on any feelings of empathy towards any members of the enemy group. It seems that at least some military leaders realize and acknowledge that war crimes are its inevitable by-product. For instance, Gonzalez refer to a commander who reported that:

I've seen my own men commit atrocities, and should expect to see it again. You can't stimulate and let loose the animal in the man and then expect to be able to cage it up again at a moment's notice.¹⁴⁶

Other leaders feel that “atrocities encourage offensive spirit” and that “punishing such behavior would likely reduce appropriate aggressive behavior in battle.”¹⁴⁷

However, at the moment, those in charge of the military – the high-ranking officers and commanders who have the power to influence what kind of training their soldiers receive – have virtually no incentive to re-examine their practices and ensure that training and socialization in the military is geared towards preventing atrocities. It seems to be advantageous for them to cultivate blanket aggression towards the enemy in their soldiers, and they might believe that a more nuanced approach in their indoctrination – for instance, including behavioral training in which soldiers practice refusing illegal orders – would undermine the effectiveness of their armies. Further, they are shielded entirely from responsibility when the soldiers they have indoctrinated to kill – without also providing effective training to make certain they do not kill indiscriminately – are unleashed in the course of conflict, provided that they do not actually intend or know about any specific atrocities that some of their soldiers will almost inevitably commit as a result of being deployed to secure the military's goals with violence.

It is arguably much more just for the law to demand that those who benefit most from warfare should bear the primary responsibility for and pay the price for the consequences of that warfare, than for that primary responsibility to remain on the shoulders of those who have little real control over their actions in the chaos of war. Further, it is also more realistic to expect those who design, implement and approve military training would be able to make rational decisions in relation to that training, taking legal requirements into account. This is in particular because such decisions would normally be administrative measures outside the context of

¹⁴⁶ Gonzalez (2004) at pp.133-134

¹⁴⁷ *Ibid*

any specific conflict – thus divorced from the emotional and social pressures operating on the minds of those already fully engaged in a group conflict. In this way, the responsibility for prevention of atrocities is placed on much more rational decision makers higher up the chain of command – people whose decision making processes are less impaired, and who can (and should) be expected to familiarize themselves with the root causes of human behavior that leads to atrocities in war, and seek to address those causes. In light of the foregoing, it therefore seems reasonable for the law to require that those most responsible for training and then unleashing a destructive force to achieve their ends should be responsible for any unintended side-effects – particularly when “unintended” does not mean “unforeseeable” or “unpreventable,” and the “side-effects” can be as grave as serious war crimes, genocide or crimes against humanity.

It should be possible to expand the notion of command responsibility to accommodate liability for this kind of failure to ensure adequate training of one’s combatants (and/or deploying such improperly trained combatants into a conflict), which would crystalize when one or some of them committed an international crime. The duty to take all necessary and reasonable measures prevent crimes is already part of command responsibility under Article 28(a)(ii), and The Pre-Trial Chamber in *Bemba* confirmed that an obligation to ensure that soldiers receive adequate training in international humanitarian law is part of that duty.¹⁴⁸ The Pre-Trial Chamber has also indicated that, while an element of causation is required, there is no need to show a direct causal link between the superior’s failure to take such measures and the crime committed by his subordinates – only that said failure increased the risk of the commission of crimes charged.¹⁴⁹ Given the results of behavioral research discussed above, it is clearly arguable that the current indoctrination and socialization of soldiers increases the risk they will commit crime; it therefore follows that failing to adjust the humanitarian training that soldiers receive to compensate for the

¹⁴⁸ Decision on the Charges, *Prosecutor v Jean-Pierre Bemba Gombo*, (ICC-01/05-01/08), Pre-Trial Chamber II, 15 June 2009 at para 437; unfortunately, what constitutes “adequate” training does not appear to include any behavioral training or socialization at present.

¹⁴⁹ *Ibid* at para 425

effects of in-group bias and social pressures can be seen as a violation of the Article 28(a)(ii) duty to prevent crimes from being committed.

Further, in the same decision, the ICC indicated that it interprets the “should have known” constructive knowledge standard as amounting to negligence.¹⁵⁰ Commanders are therefore already liable for any unreasonable failure to exercise supervision by seeking out relevant information about what their troops are doing in the field, where that failure prevents them from discovering, deterring and punishing international crimes – even, it appears, in absence of any notice that criminal activity is actually happening. It does not seem to be a great departure from the law as it is at present to include under this head of liability the liability for negligence on the part of the top military command in failing to deter their soldiers from committing crimes by instituting appropriate training, in circumstances where they ought to have known that such failure will result in some of their troops committing crimes. Given the serious nature of the crimes in question – and the correspondingly large public interest in deterring such crimes – there may even be an argument for extending the law a step further, to impose a principle that those responsible for the training and socialization of combatants are presumed to be liable for any crimes those combatants commit in the course of conflict. Such a presumption could be rebutted by evidence that adequate training was provided to ensure that soldiers react appropriately, or that the crime was unrelated to his socialization or training.

However, there appears to be a temporal limitation to the liability that can be imposed under Article 28 – superiors are only liable when they know (actually or constructively) of crimes their subordinates are committing or about to commit (Article 28(a)(i)). This would seem to eliminate responsibility for possible future crimes, committed as a foreseeable result of poor training. Further, the expanded notion of constructive liability accepted by the Pre-Trial Chamber in *Bemba*, which includes negligence, only applies to military commanders – civilian leaders, who might have as much say in how combatants are trained, are not liable

¹⁵⁰ *Ibid* at paras 429 and 433; see discussion at pp.21-22 above.

unless they either have actual knowledge of the crime being committed or about to be committed, or “consciously disregarded information which clearly indicated, that the subordinates were committing or about to commit such crimes” (Article 28(b)(i)). To impose primary liability for crime on those ultimately responsible for the training and socialization of soldiers, and thus incentivize such people to provide training likely to deter their subordinates from committing crimes, both of these requirements would have to be significantly revised (or interpreted very generously by the ICC).

5.4 Potential problems with allocating liability as proposed

5.4.1 Cognitive error in law making

Instituting changes in established law generally requires that the law makers – the judges and legislators who write and interpret the law – have a sufficiently good understanding of the necessity for change, and the will to implement it. Unfortunately, to alter the allocation of international criminal responsibility in the manner proposed would require judges and legislators to adjust quite dramatically their understanding of human behavior, and the importance of one’s society and circumstances in the commission of international crimes. The problem is not only that judges lack the relevant knowledge of psychology, but also that they already have a firmly established view of human behavior that they believe to be correct, but which conflicts with psychological evidence. As the people who create our laws are as human as the rest of us, their minds are prone to the same cognitive errors that everyone experiences – including confirmation bias and cognitive dissonance. This means that they may be prone to dismiss evidence that conflicts with their long-established ideas about “moral choice” and “reasoning agents,” or try to re-interpret it in such a way that would allow them to still maintain a belief in the little Captain Rationality at the center of our thinking. Or they may do what the judges in *Erdemović* did and refuse to consider it at all – the Appeal Chamber rejected an application to introduce additional evidence about the possibility of “moral choice of an ordinary soldier who is faced with committing a crime when

following the orders of a superior”¹⁵¹ as well as an application for a further psychological examination concerning Erdemović’s state of mind at the time he committed the offence.¹⁵²

Further, most of us – including judges - also suffer from something that can be called the illusion of competence (also known as the “better-than-average” effect). This can be described as the belief that our own performance and ability is better than average on most common skills and characteristics. For instance, the majority of people will rate themselves as closer to “expert” than “poor” in driving skills, even when they are asked to provide such rating when they are in a hospital recovering from an automobile accident.¹⁵³ Combined with naïve realism (the belief that reality is exactly as we perceive it – that we see the world as it truly is, rather than as it is interpreted by the complicated perceptual and cognitive machinery that is our brain),¹⁵⁴ the faith in our own competence leads us to assume that we personally would be less affected than other people by bias, pressures to conform, and other influences. This illusion persists even after we have been shown evidence that clearly indicates that the majority of people succumb to such influences. As one prominent psychologist put it:

The brain is designed with blind spots, optical and psychological, and one of its cleverest tricks is to confer on us the comforting delusion that we, personally, do not have any.¹⁵⁵

The belief that we can resist influence that affects others, leads us to believe that such resistance is a realistic option for others; in turn, this makes us less sympathetic to those lesser mortals who succumb to such influence. Judges in international courts very likely cannot bring

¹⁵¹ Judgment, *Prosecutor v Dražen Erdemović* (IT-96-22-A), Appeals Chamber, 7 October 1997 at paras 14 and 15. The application was for such evidence to be presented by a “professor of ethics” rather than a psychologist; nevertheless, the topic should have been relevant to their determination of whether Erdemović was capable of moral choice at the crucial time – that they found otherwise on the basis that this evidence was irrelevant, indicates that they must have had established ideas of what “moral choice” is, and when it is possible to make it, and were perhaps not interested in having those beliefs challenged.

¹⁵² *Ibid.*

¹⁵³ Gilovich et al (2016) at p.91

¹⁵⁴ *Ibid.*, at pp.14-15. This belief is of course false – the reality we consciously perceive is that constructed by our brains, based on both inside and outside input.

¹⁵⁵ Tavis and Aronson (2007) at p.42

themselves to accept that they could ever act in the manner that the defendants appearing before them have acted – even if they understand about the power of social and circumstantial pressures, their belief in their own competence (and thus the ability to overcome such pressures) will lead them to prefer to see the criminal act in question in terms of the defendant’s weakness of character rather than the power of the situation that the defendant found himself in.

What most judges probably fail to understand or internalize is that it is very probable – one might even go as far as to say, overwhelmingly likely – that given the right set of circumstances, they too would commit atrocities. This conclusion can be drawn not only from the experimental evidence of Milgram and others, but also from the work of such people as Hans Petter Graver who, having looked at the track record of the judicial action under various oppressive regimes, found a shameful history of content, or even eager, enforcement of unjust, tyrannical and murderous laws by the vast majority of judges, in violation of basic human rights and elementary morality.¹⁵⁶

5.4.2 What would constitute effective training?

The second major problem is that there is thus far no evidence specific to the combat setting that altering the way that soldiers are trained to behave in accordance with humanitarian law would have any deterrent or preventive effect on international crime. An associated problem is how to ascertain what would constitute “effective training” for the purposes of either rebutting the proposed presumption or fulfilling the duty to prevent crime under Article 28 in its expanded meaning.

One way to answer such objections is to point out that there is already a duty to provide training in humanitarian law to combatants, despite the fact that such training does not appear to do much at present to prevent atrocities. The proposed alteration in attribution of liability will hopefully provide sufficient incentive for those responsible for such training

¹⁵⁶ H. P. Graver (2015), *Judges Against Justice: On Judges When the Rule of Law is Under Attack*, Springer-Verlag, Berlin and Heidelberg

to find methods that do work and are effective at increasing deterrence from crime, based on a much firmer ground of actual research into human behavior. Much of the military training already relies on behavioral research to ensure that their soldiers kill enemies when ordered to do so.¹⁵⁷ This training is as effective as it is because it relies on shaping each soldier's instincts, rather than appealing to any reasoning process – the soldiers are not just told why they must obey and kill, they are made to practice it over and over again in various circumstances until it becomes a near-automatic reaction. There is no reason why the military should not be able to come up with equally effective behavioral training that would ensure soldiers comply with the basic rules of humanitarian law. Such training would likely have to include repeated practice in how to recognize and disobey illegal orders and resist social pressure, as well as letting soldiers experience, and reflect upon, bad decision-making. Educating soldiers on the way that their decision-making process works, and warning them about cognitive errors might also be quite useful, as would instituting standard procedures to prevent groupthink.¹⁵⁸ Finally, social indoctrination should emphasize humanizing civilians (including enemy civilians). This could be done by reading accounts of victims of war crimes and reflecting upon those accounts; or encouraging heroic tales about saving civilians by respected group members. On the other hand, dehumanizing language or practices should be strictly penalized; if it is impossible to eliminate such behaviors entirely (given that the ability to kill might rely, at least partly, on dehumanizing the person to be killed), some effort should be made to target dehumanization more precisely at a very narrow category of enemies (i.e. “people who are armed and currently trying to kill us”), always excluding civilians and prisoners of war.

¹⁵⁷ Gonzalez (2004) at p.119-121

¹⁵⁸ Groupthink is “a kind of faulty thinking by highly cohesive groups in which critical scrutiny that should be devoted to the issues at hand is subverted by social pressures to reach consensus.” (Gilovich et al at p.457) The Bay of Pigs incident and the decision not to take precautions at Pearl Harbor despite warnings of an imminent attack, are both particularly disastrous results of groupthink. For more on groupthink and how to effectively prevent it, see: Gilovich et al (2016) at p.457-460

5.4.3 The physical perpetrators

Another problem is what to do with the perpetrators of the crimes committed under social and circumstantial pressures, if primary liability for such crimes is to be shifted onto the people in high command responsible for their training. If the legal system accepts social pressure (combined with the absence of proper training to resist said pressure) as a valid defence to their crimes, this does not necessarily mean that such perpetrators should be released. In light of the research set out above, it is likely that most of the perpetrators of international crimes will view the atrocities they committed as justified (or at least excused), and maintain a dehumanized view of their victims and other members of the enemy group to which the victims belonged. Erdemović's sincerely remorseful attitude seems to be quite rare – to truly repent of one's actions requires a person not only to accept that his actions were wrong (or even monstrous), but also to empathize with his own victims. As the findings on cognitive dissonance and morality set out above should make clear, this is not something that is terribly likely to happen naturally – most of us will seek to justify or excuse our actions, and shift the blame onto the victims.

Even though we might understand that such an unrepentant attitude is a product of our natural cognitive process (which we cannot fully control), that may be little to no comfort to the surviving victims and relatives of victims of such perpetrators, who might feel freshly traumatized by what they would perceive as a gross injustice of their release. Further, such perpetrators would then return – unchanged in their attitude – to their communities, where they would likely continue reinforcing and spreading hatred against the enemy group. Even if inter-group violence never ignites again, that is surely not a desirable result, given a multitude of undesirable behaviors that can take permanent root in such communities, such as discrimination and individual attacks against members of the enemy group.

On the other hand, imprisonment seems both unjust (given the circumstances under which their crimes were committed) and can have a somewhat counter-productive effect on the perpetrator's society, given that societies involved in a conflict (past or present) have a disturbing tendency

to view their war criminals as martyred heroes. This is what appears to have happened, for instance, in the case of Balkan war criminals prosecuted and/or convicted by the ICTY: a 2009 OSCE survey found, *inter alia*, that over half of their Serbian respondents thought that the evidence against Serbian war criminals was false;¹⁵⁹ and recently, a convicted war criminal (Vladimir Lazarevic) was flown home by two senior Serbian ministers and given a hero's welcome in Serbia after being released early from his sentence.¹⁶⁰

One option may be to pursue re-education of the perpetrator, combined with reconciliation with his victims (if possible) and some kind of community service. The goal would be to re-humanize the victims to the perpetrator, get him to feel remorse for the acts he had committed, and allow him to offer reparation to the victims or the victims' community. This should make it less likely that the perpetrator would engage in any further discriminatory or violent behavior (or language) against members of the enemy group. Once returned to his own society, the reformed perpetrators might even contribute towards a peaceful reconciliation between the two groups. Restorative justice programs can also offer much better results for victims, who report "fewer thoughts of revenge than comparison individuals," and who are also "more than twice as likely to forgive the offender and say that the criminal justice system is fair."¹⁶¹

5.4.4 Defining the guilty

An additional difficulty might arise with defining the parties upon whom liability ought to be imposed, as a wide array of actors may play a part in putting an improperly trained combatant into the field of conflict. They can include professionals responsible for designing and implementing the training that the soldier receives (psychologists and social scientists, as

¹⁵⁹ OSCE (2009), *Views on war crimes, the ICTY, and the national war crimes judiciary*, Strategic Pulse Group, found on <http://www.osce.org/serbia/40751?download=true>, accessed on 26 May 2016

¹⁶⁰ M. Ristic (3 Dec 15), "Serbia Welcomes Freed Yugoslav Army War Criminal," *Balkan Transitional Justice*, found at <http://www.balkaninsight.com/en/article/yugoslav-army-war-criminal-arrives-to-serbia-12-03-2015>, accessed on 26 May 2016

¹⁶¹ Gilovich et al (2016) at p.518.

well as professional military trainers), generals or some other commanders who commission said training and/or deploy poorly trained soldiers into the field, or members of any oversight committee (whether military or civilian) charged with monitoring or funding the training given. But the net could potentially be cast even wider to include field commanders who fail to inquire into the kind of training that the soldiers they accept into their units receive; on the other hand, such commanders (particularly if already embroiled in a conflict), could be subject to the same group pressures that affect their subordinates and too low in the military hierarchy to have any real say over deployment. Further work thus needs to be done to examine the relative power, influence and capacity for rational thought of all of these (and any other) potential actors, in order to more precisely target those who are both responsible and deterrable.

6. Conclusion

The way that liability for international crimes is currently allocated – concentrating on the physical perpetrators and those directly involved in the crime – mostly ignores the complex causal roots of the criminal behavior, and the importance of social and circumstantial factors in the commission of such crimes. Instead, the law continues to examine human behavior in terms of rational choices made by reasonable individuals, clinging to an unrealistic “homunculus” model of human thought, in support of a narrow and antiquated version of individual responsibility that takes no account of the fact that as a social species, our society plays a large causative role in all of our behaviors.

If international criminal law is to have any effective role in preventing and deterring international crimes, it must ground its rules in the evidence of how humans actually behave. This means that it must strive to allocate criminal liability for international crime to people who both have the power to institute effective crime-prevention programs (such as appropriate training for combat personnel) and are in the position to realistically take legal rules into account in their decision making process. This requires judges and legislators to re-examine in good faith their basic assumptions about human behavior – including such traditional notions as “moral choice” and the existence of “reasoning agents” on the battlefield – and take into account what the discipline that actually studies human behavior can tell us about what effective deterrence consists of. Or as one (clearly somewhat frustrated) scholar put it:

The workings of the criminal justice system should not continue to be guided by illusions about cross-situational consistency in behavior, by erroneous notions about the impact of dispositions versus situations in guiding behavior, or by failures to think through the logic of person x situation interactions, or even comforting but largely fanciful notions of free will, any more than it should be guided by once common notions about witchcraft and demonic possession.¹⁶²

¹⁶² L. Ross and D. Shestowsky (2003), “Symposium: Empirical Legal Realism: A New Social Scientific Assessment of Law and Human Behavior: Contemporary Psychology’s

Challenges to Legal Theory and Practices, *Northwestern University Law Review*, 97, at pp.1081-1114

Bibliography

Bohrer, Z (2012), "Is the Prosecution of War Crimes Just and Effective? Rethinking the Lessons from Sociology and Psychology," *Michigan Journal of International Law*, Vol.33, pp.749- 819

Cassese, A and Gaeta, P (2013) *International Criminal Law, 3rd Ed*, Oxford: Oxford University Press

Chantal, M (2007) "Command Responsibility: Mode of Liability for the Crimes of Subordinates or Separate Offence of the Superior," *J Int Criminal Justice*, Vol. 5(3), pp.619-637

Cushman, F., Mendes, W.B., et al. (2012) "Brief Report: Simulating Murder: The Aversion to Harmful Action." *Emotion*, Vol 12, No 1, pp.2-7

Danziger, S, Levay, J and Avnaim-Pesso, L (2011), "Extraneous factors in judicial decisions," *Proceedings of the National Academy of Sciences of the United States*

Eysenck, M. W, and Keane, M. T (2015) *Cognitive Psychology: A Student's Handbook, 7th Ed*, London and New York: Psychology Press

Festinger, L et al (2008) *When Prophecy Fails*, London: Pinter & Martin

Gaeta, Paola (1999) "The defence of superior orders: The Statute of the International Criminal Court *versus* customary international law", *European Journal of International Law*, Vol. 10, pp. 172-191

Gilovich, T et al (2016), *Social Psychology, 4th Ed*, New York and London: W. W. Norton and Company

Goodenough, O. R. (2006) "Can Cognitive Neuroscience Make Psychology a Foundational Discipline for the Study of Law?" in *Law and Psychology: Current Legal Issues Volume 9*, Brooks-Gordon, B and Freeman, M eds. Oxford: Oxford University Press

Goodenough, O. R. and Tucker, M (2011), "Neuroscience Basics for Lawyers," *Mercer Law Review*, 62(3), pp.945-958

Gonzalez, N (2004), *Moral Monsters or Ordinary Men Who Do Monstrous Things? Psychological Dimension of the Military and Their Implications for War Crimes Tribunal Defences*, unpublished PhD thesis, Widener University.

Graver, Hans Petter (2015), *Judges Against Justice: On Judges When the Rule of Law is Under Attack*, Springer-Verlag, Berlin and Heidelberg

Greene, J. (2013) *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them* (Kindle edition) London: Atlantic Books

Green, L.C. (1976) *Superior orders in national and international law*, Leyden: A. W. Sijthoff

Hamlin, J. K., Wynn, K., Bloom, P. (2007) "Social evaluation by preverbal infants." *Nature* 450, pp.557-559

Kahneman, D (2011), *Thinking Fast and Slow*, London: Penguin Books

Kapardis, A (2010) *Psychology and the Law: A Critical Introduction*, 3rd Ed, Cambridge: Cambridge University Press

Knoops, G-J. A. (2008) *Defences in Contemporary International Criminal Law*, 2nd Ed, Leiden: Martinus Nijhoff Publishers

Latané, B and Darley, J. M (1968), "Group Inhibition of Bystander Intervention in Emergencies," *Journal of Personality and Social Psychology*, 10(3), pp.215-221

Milgram, S (1963), "Behavioral study of obedience," *Journal of Abnormal and Social Psychology* 67, pp.371-378

Milgram, S and Blass, T eds.,(2010) *The Individual in a Social World: Essays and Experiments*, 3rd ed., Pinter & Martin, London

Ohlin, J (2009) "Necessity and Duress," in Cassese, A, *The Oxford Companion to International Criminal Justice*, Oxford: Oxford University Press

Oliner, S and Oliner, P (1988), *The altruistic personality*, New York: Free Press

OSCE (2009), *Views on war crimes, the ICTY, and the national war crimes judiciary*, Strategic Pulse Group

Osiel, M (2001), *Mass Atrocity, Ordinary Evil and Hanna Arendt: Criminal Consciousness in Argentina's Dirty War*, Yale University Press

Osiel, M (2005) "The Banality of Good: Aligning Incentives Against Mass Atrocity," *105 Columbia Law Review*, pp.1751-1862

Penny, C. K. (2010) "Obeying Restraints: Applying the Plea of Superior Orders to Military Defendants before the International Criminal Court," *Canadian Yearbook of International Law*, Vol. 48, pp.3-38

Ristic, M (3 Dec 15), "Serbia Welcomes Freed Yugoslav Army War Criminal," *Balkan Transitional Justice*

Roberts, D.L. (1997) "Training the armed forces to respect international humanitarian law: The perspective of the ICRC Delegate to the Armed and Security Forces of South Asia," *International Review of the Red Cross*, No.319

Ross, L and Shestowsky, D (2003), "Symposium: Empirical Legal Realism: A New Social Scientific Assessment of Law and Human Behavior: Contemporary Psychology's Challenges to Legal Theory and Practices," *Northwestern University Law Review*, 97, at pp.1081-1114

Tavris, C and Aronson, E (2007), *Mistakes were made (but not by me): Why we justify foolish beliefs, bad decisions and hurtful acts*, London: Pinter and Martin

Taylor, K (2004), *Brainwashing: The science of thought control*, Oxford: Oxford University Press

Taylor, K (2009), *Cruelty: Human Evil and the Human Brain*, Oxford: Oxford University Press

Weingend, Thomas (2012) "Kill or Be Killed: Another Look at Erdemović," 10 *J. Int'l Crim. Just.* 1219-1237

Werle, G (2007) "Individual Criminal Responsibility in Article 25 ICC Statute," *Journal of International Criminal Justice* 5(4), pp. 953-976

Young, L and Koenigs, M (2007), "Investigating emotion in moral cognition: a review of evidence from functional neuroimaging and neuropsychology." *British Medical Bulletin*, Vol.84, pp.69-79

Zajonc, R.B. (1980). "Feeling and Thinking: Preferences Need No Inference." *American Psychologist*, Vol. 35, No.2, 151-175

Zimbardo, P (2007) *The Lucifer Effect: How Good People Turn Evil*, Random House, London

Table of Cases

Decision on the Charges, *Prosecutor v Jean-Pierre Bemba Gombo*, (ICC-01/05-01/08), Pre-Trial Chamber II, 15 June 2009

Judgment, *Prosecutor v Jean-Pierre Bemba Gombo*, (ICC-01/05-01/08), Trial Chamber III, 21 March 2016

Re Buck (1946) 13 Ann. Dig. 293

Caroelli and others, Italian Court of Cassation 10 May 1947

Judgment, *Delalić and others*, (IT-96-21-A), Appeals Chamber, 20 February 2001

Sentencing judgment, *Erdemović* (IT-96-22-T), Trial Chamber, 29 November 1996

Judgment, *Prosecutor v Dražen Erdemović* (IT-96-22-A), Appeals Chamber, 7 October 1997

Separate and Dissenting Opinion of Judge Cassese, *Prosecutor v Dražen Erdemović* (IT-96-22A), Appeals Chamber, 7 October 1997

Joint Separate Opinion of Judge McDonald and Judge Vohrah, *Erdemović* (IT-96-22-A), 7 October 1997

Sentencing Judgment, *Erdemović* (IT-96-22-Tbis), Trial Chamber, 5 March 1998

Trial of Gustav Alfred Jepsen and Others, Proceedings of a War Crimes Trial held at Lüneburg, Germany (13-23 Aug. 1946), Judgement of 24 Aug. 1946

In re List and Others (Hostages Trial), Judgment 215 (19 February 1948)(USMT, Nuremberg), 15 ILR 632

Trial of Otto Ohlendorf et al., (“*Einsatzgruppen*” case), in *Trials of War Criminals before the Nuernberg Military Tribunals under Control Council Law No. 10*, (U.S. Govt Printing Office, Washington D.C., 1950)

Parker v Levy, 417 U.S. 733 (1974)

Stenger and Crusius, Leipzig Supreme Court, 6 July 1921, *Verhandlungen*, 2563-72

Judgment, *Strugar*, (IT-01-42-A), Appeals Chamber, 17 July 2008

Judgment, *Tadić* (IT-94-1-A), Appeals Chamber, 15 July 1999