cognitio

studentisches Forum für Recht und Gesellschaft

From Is to Ought

How Scientific Research in the Field of Moral Cognition Can Impact the Criminal Law*

LEVIN S. GÜVER[1]

Rapid technological advances in the field of neuroscience and cognitive psychology are claiming to have solved the millennia-old puzzle of moral cognition. If true, our societal structures – and with that the criminal law – would be gravely impacted. This paper concerns itself with four distinct theories stemming from the disciplines above, taking an in-depth look at the Dual Process Theory by JOSHUA GREENE and juxtaposing the findings to the consequentialist and retributivist theories of punishment present in the American Criminal Law Doctrine.

Table of Contents

I. Preface 2
A. Historical Classification 2
B. Terminology and Methodology 3
II. Cognitive Account 4
A. The Stage Model 4
B. Requirements for Moral Development 5
C. The Stages of Moral Development 5
D. Mentalism and the Universal Moral Grammar Theory 6
III. Emotive Account 7
A. Moral Dumbfounding 7
B. The Social Intuitionist Model 7
IV. Dualistic Account 8
A. Approach 8
B. Results 9
C. The Dual-Process Theory 10
V. Foundation of the Criminal Law 10
A. Purpose 10
B. Theories of Punishment 11
C. The Normative Impact of Science 11
VI. Direct Approach to Rejecting Retributivism 12
A. Linking Emotion and Deontology 12
B. Linking Deontology and Retributivism 14
C. The Deontic Core of the Mental Gizmo Thesis 14
D. Deontology as Heuristics 15
E. Interim Conclusion 15
VII. Indirect Approach to Rejecting Retributivism 16

* This paper was initially submitted as a Bachelor Thesis for a seminar by Prof. Matthias Mahlmann (UZH) and Prof. Stephen J. Morse (UPenn) in December 2018, Zürich. I am indebted to Famke for her continuous support, without which I could not be here.

1 Levin Sinan Güver, BLaw (University of Zurich); levin.guever@gmail.com.

A. Presupposed Picture of the Law 16
B. The Problem of Free Will 16
C. The Curious Case of Mr. Puppet 17
1. Approaching from Neuroscience 17
2. Approaching from Physics 18
D. The Curious Case of Mr. Puppet Continued 18
VIII. Conclusive Remarks 19
IX. Appendix 21

I. Preface

«In every system of morality, which I have hitherto met with, I have always remark'd, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surprizʼd to find, that instead of the usual copulations of propositions, is, and is not, I meet with no proposition that is not connected with an ought, or an ought not. This change is imperceptible; but is, however, of the last consequence. For as this ought, or ought not, expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it.»[2]

A. Historical Classification

Views on moral cognition have undergone several changes throughout the course of history, the conflict between reason and emotion being firmly rooted in its core. Ancient philosophers deemed it «a conflict between divinity and animality».[3] SOKRATES traced the origin of moral judgement back to the *daimonion*, an inner voice guiding our sense of morality.[4] In the 18th century, DAVID HUME questioned rationalism and swung the pendulum in favor of an emotivistic approach,[5] claiming moral judgement to be the result of «some internal sense or feeling, which nature has made universal to the whole species».[6] In an attempt to refute HUME,[7] IMMANUEL KANT created his *rationalist ethical theory*, arguing moral judgement to be the result of practical reason guided by the categorical imperative.[8]

This domain – once exclusive to what we would now deem «armchair philosophers»[9] –

2 HUME DAVID, A Treatise of Human Nature, 1739, L. A. Selby-Biggie (ed.), 1896, Book III, Part I, Section I.

3 HAIDT JONATHAN, The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgement, in: Psychological Review, vol. 108, no. 4, 2001, p. 815.

4 PLATO, Apology, F. J. Church (trans.), 1963, 31 c, d.

5 Maintaining the basic position that «the ultimate ends of human actions can never [...] be accounted for by reason, but recommend themselves entirely to the sentiments and affections of mankind», HUME DAVID, An Enquiry Concerning the Principles of Morals, 1777, prod. J. Mamoun, C. Franks, 2010, Appendix I, Section V.

6 HUME (Fn. 5), Appendix I, Section I.

7 Despite their paradigmatic differences, KANT held HUME in high regard, going as far as saying HUME awakened him from his «dogmatic slumber», KANT IMMANUEL, Prolegomena to Any Future Metaphysics, 1783, G. Hatfield (trans. a. ed.), 2004, XIV.

8 One could have a far-ranging discussion on whether KANT *created* or *discovered* the categorical imperative. For its three versions, cf. KANT IMMANUEL, Grundlegung zur Metaphysik der Sitten, 1785, in: I. Kant, Akademieausgabe von Immanuel Kants Gesammelten Werken, Presussische Akademie der Wissenschaften (ed.), vol. IV, 1911, p. 421, 429, 437.

9 Courtesy of MAHLMANN MATTHIAS, Mind and Rights: Neuroscience, Philosophy, and the Foundations of Legal Justice, in: M. N. S. Sellers (ed.), Law, Reason and Emotion, 2017, p. 121 fn. 151. While one might argue that the term cannot be applied to philosophers before certain technological advances – claiming philosophy was a purely mental domain due to the inability to conduct wide-scale research –, it is important to keep in mind that studies are just one of many approaches towards empirical evidence and theories have been refined through observation since ancient times. One can still ponder whether the great minds of the past would have made use of the tools we have at our disposal today (for example fMRI scans) or if they would have regarded certain questions and domains to be

became part of the multidisciplinary empirical movement in the late 19th century, urging psychologists to «abandon their armchairs and go into the laboratories» in an attempt to unravel the mysteries of the human mind.[10] Rapid scientific progress gave rise to neuroscience, which claims to have captured the problem of morality at its core – the human brain.[11] But is neuroscience really capable of opening up the black box of cognition?

The following paper will be devoted to apprehending the current state of psychological and neuroscientific research on moral cognition and applying the insights to the American criminal law doctrine, structured in style of the *Is-Ought Problem* HUME famously postulated.[12] Commencing with a brief definition of the topics at hand, it will narrow down to four distinct takes on moral judgement, providing a comprehensive account of the Dual-Process Theory in particular – a blend between neo-emotivistic[13] and cognitivistic approaches. The findings will be juxtaposed against foundational notions of the criminal law, questioning its underlying principles and highlighting the current cleft between what is and ought to be.

B. Terminology and Methodology

When it comes to defining the terms *morality* and *moral* judgement, there is – as with most topics up for philosophical debate – no distinguished meta-definition. However, it is possible to identify certain reoccurring elements and attempt to construct one accordingly. Such conceivable attempt may look like this:

Morality is a [universal] system of principles and values, distinguishing between good and bad acts. Moral cognition is the individual's ability to tap into that system, constituting moral judgement and internally nudging the individual towards the moral act.[14]

exempt of the scientific method, surrendering only to the efforts of the mind.

10 HAIDT (Fn. 3), p. 816. It would be naive to think that the research was purely empirical with no prior theoretical considerations. E.g., KOHLBERG recognized the significant relationship between theory and empiricism, labelling it a «spiral circularity», KOHLBERG LAWRENCE, A Current Statement on some Theoretical Issues, in: S. & C. Mogdil (eds.), Lawrence Kohlberg. Consensus and Controversy, 1986, p. 505. GARZ supplies an interesting comparison, highlighting its similarity to a shoelace, «which is made up of one piece, then separated and pulled apart at the beginning of the threading only to be rejoined later», GARZ DETLEF, Lawrence Kohlberg – An Introduction, 2009, p. 30.

11 Quoting MICHAEL GAZZANIGA, «98 or 99 percent» of cognitive neuroscientists support the reduction of the mind to the brain, cf. SNEAD CARTER O., Neuroimaging and the «Complexity» of Capital Punishment, in: New York University Law Review, vol. 82, no. 5, 2007, p. 1279. Taken one step further, *eliminative materialism* seeks to reduce human action and behavior to corresponding brain states. For a short overview, cf. LELLING ANDREW E., Eliminative Materialism, Neuroscience and the Criminal Law, in: University of Pennsylvania Law Review, vol. 141, no. 4, 1993, p. 1476; PARDO MICHAEL S., PATTERSON DENNIS, Philosophical Foundations of Law and Neuroscience, in: University of Illinois Law Review, vol. 2010, no. 4, 2010, p. 1245 f.

12 Crossing HUME'S gap is the central hardship any scientific theory has to overcome on its journey to impacting philosophy and ethics. The original quote is displayed below the preface.

13 To borrow a term from MAHLMANN MATTHIAS, Ethics, Law and the Challenge of Cognitive Science, in: German Law Journal, vol. 8, no. 6, 2007, p. 577.

14 This definition is constructed intentionally broad and incorporates elements from several authors, cf. MAHLMANN MATTHIAS, Rechtsphilosophie und Rechtstheorie, vol. 4, 2017, p. 286; GRAF TILMAN, Ethik und Moral im Grundgesetz: Grenzen der Moralisierung des Verfassungsrechts, vol. 285, 2017, p. 43 f.; GERT BERNARD/GERT JOSHUA, The Definition of Morality, in: E. N. Zalta (ed.), The Stanford Encyclopedia of Philosophy, 2017; TURIEL ELLIOT, The Development of Social Knowledge: Morality and Convention, 1983. HAIDT, in contrast, defines moral judgement as «evaluations (good versus bad) of the actions or character of a person that are made with respect to a set of virtues held by a culture or subculture to be obligatory», HAIDT (Fn. 3), p. 817. This view does not presuppose the existence of absolute moral values and only bases itself on what we unarguably know: their existence in our minds and influence on our behavior. For a similar view, see TASSY S./LE COZ P./WICKER B., Current Knowledge in Moral Cognition can Improve Medical Ethics, in: Journal of Medical Ethics, vol. 34, no. 9, 2008, p. 679. It should be noted that the idea of

While morality constitutes the system on a normative scale, moral cognition and moral judgement can be regarded as the intricate processes that take place in the individual's mind. These processes are not limited to *a priori* reasoning and fall under the scrutiny of the scientific method.[15] The inquiry process is divisible into two stages. First, the subjects are supplied with moral stimulus through confrontation with moral dilemmas[16] – situations in which opposing duties are pitted against each other, making it impossible to adhere to one without neglecting the other.[17] The second stage aims to capture the moral response in a format susceptible to further evaluation. This is where the sentiments start to diverge: whilst the psychologist might focus on behavioral aspects, the neuroscientist would consider brain scans to be the decisive piece of evidence.[18] A closer look at four conceptually distinct theories is warranted.[19]

II. Cognitive Account

A. The Stage Model

LAWRENCE KOHLBERG[20] sought to construct a cognitive-developmental framework based on prior works of JEAN PIAGET, combining academic psychology with sociology, philosophy and anthropology.[21]

He conducted moral judgement interviews in form of longitudinal studies spanning seventeen years with (initially) eighty-four boys ranging from age ten to sixteen from different socio-economic groups, confronting them with distinct moral dilemmas.[22] One of these renowned moral dilemmas is labelled the *Heinz Dilemma*:

Heinz's wife is dying from a rare form of cancer. A local druggist discovers the only known cure, but charges Heinz more than he can afford. After ex-

subjective, individually manifesting morality does not necessarily lead to a non-cognitivist view of moral relativism; the alternatives will be evident in light of the mentalist and universal moral grammar theory in section II.D, cf. MAHLMANN MATTHIAS, The Cognitive Foundations of Law – An Introduction to the Mentalist Theory of Ethics and Law, in: H. Rottleuhner, Foundations of Law, A Treatise of Legal Philosophy and General Jurisprudence, repr. edt., vol. 2, 2007, p. 76.

15 The scientific method regards the empirical side of the coin, concerning itself with the process of acquiring information through observation and experimentation. Empirical claims are used to describe the *is*, as in the observable reality of the situation, while normative claims concern themselves with how things should be, the *ought*, cf. PARDO/PATTERSON (Fn. 11), p. 1220 f.

16 See SUHLER CHRISTOPHER, CHURCHLAND PATRICIA, The Neurobiological Basis of Morality, in: J. Illes, B. J. Sahakian (eds.), The Oxford Handbook of Neuroethics, 2011, p. 34 f.

17 See ELSIGAN ALFRED, Gibt es «echte» moralische Dilemmata? Das Trolley-Problem, 2014, p. 13.

18 The most common method being functional Magnetic Resonance Imaging (fMRI). Neuronal activity and cerebral blood flow are coupled, making it possible to deduce which brain areas are in use through a procedure based on the magnetic difference between oxygenated and deoxygenated blood, cf. OWEN ADRIAN M., Functional Magnetic Resonance Imaging, Covert Awareness, and Brain Injury, in: J. Illes, B. J. Sahakian (eds.), The Oxford Handbook of Neuroethics, 2011, p. 137; SCHLEIM STEPHAN, Über einen möglichen normativen Beitrag der Moralphysiologie, in: G. Sharifi (ed.), Brauchen wir eine neue Moral?: Herausforderungen der Ethik durch die Neurowissenschaften, 2011, p. 183.

19 It has to be stressed that the following theories do not provide an exhaustive account on the current discourse of moral judgement – they are intended as a comprehensive overview.

20 LAWRENCE KOHLBERG was regarded as a leading figure in the *cognitive revolution*, cf. HAIDT (Fn. 3), p. 816.

21 Cf. GARZ (Fn. 10), p. 26; ROTTMAN JOSHUA/YOUNG LIANE, Mechanisms of Moral Development, in: J. Decety, T. Wheatley (eds.), The Moral Brain: A Multidisciplinary Perspective, 2015, p. 123.

22 Cf. GARZ (Fn. 10), p. 39 f. KOHLBERG was subject to a lot of criticism, as he assumed moral cognition to be uniform between the genders – an assumption he deeply regretted later on, cf. GARZ DETLEF, Kohlberg zur Einführung, corr. vol. 2, 2015, Appendix 9. For a further read on KOHLBERG'S gender bias, cf. WALKER LAWRENCE J., Progress and Prospects in the Psychology of Moral Development, in: Merrill-Palmer Quarterly, vol. 50, no. 4, 2004, p. 551 f.

hausting every legal means, Heinz decides to break into the store. Should Heinz steal the drug?[23]

In addition to the interviews, KOHLBERG made use of a variety of survey methods, such as evaluations from close peer groups and tasks on role-taking.[24] The interviews were dedicated to bringing forth the subject's most advanced form of reasoning, and – combined with the subsequent survey methods – sought to detect the underlying *deep structures* behind moral development.[25]

B. Requirements for Moral Development

KOHLBERG's work is centred around the view of morality as an universal justice structure,[26] concerning itself with the interdependence of rights and responsibilities.[27]

He deems three characteristics central for moral development:

(1) Innate universal social institutions such as family, economy, law, and government.[28] Adhering to these institutions requires the ability to interchange perspectives through empathizing processes, leading KOHLBERG to conclude that both society and morality are «a structure of interaction between the self and other selves who are like the self, but who are not the self».[29]

(2) Key concepts – most notably justice – differentiating between various levels of theoretical difficulty.[30]

(3) Social stimulation as the motor behind moral development, achieved by partaking in social events from institutions listed in (1).[31]

C. The Stages of Moral Development

Evaluating his results, KOHLBERG created a framework consisting of six *hard stages* – increasing in sophistication – which he deemed vital for the development of moral judgement.[32] The individual can progress

23 Cf. GARZ (Fn. 10), p. 55. The dilemma continues with alternate scenarios aiming to capture the subject's full scope of moral judgement.

24 Cf. GARZ (Fn. 10), p. 40.

25 Cf. GARZ (Fn. 10), p. 33. KOHLBERG based this on a theory of competence inspired by NOAM CHOMSKY, who differentiates between a subjects linguistic competence and performance: the competence is based on a theoretical, idealized starting point, while the performance is his actual, displayed use of language, cf. CHOMSKY NOAM, Aspects of the Theory of Syntax, 1965, p. 3 f. KOHLBERG thus did not limit himself to the subjects performance, but sought after his competence in an attempt to unravel innate *deep structures* of moral judgement. For a comprehensive overview, see GARZ (Fn. 10), p. 31 ff.

26 Regarding justice to be on the forefront of morality is not an unorthodox view. GIBBS claims an adequate morality to require »both the right and the good«, suggesting its foundation in empathy and fairness, cf. GIBBS JOHN C., Moral Development and Reality: Beyond the Theories of Kohlberg, Hoffman, and Haidt, 2015, p. 7. Similarly, MAHLMANN regards both altruism and the «justice-as-proportional-equality-principle» to constitute the foundational judgements of morality, regarding empathy as a «central heuristic tool» for moral judgement, cf. MAHLMANN (Fn. 13), p. 587, 593 ff. KOHLBERG deemed the universality aspect of special importance, conducting his studies in various parts of the world.

27 Cf. GARZ (Fn. 10), p. 8. A similar view is held by BAUMARD/SHESKIN, who regard morality as questions concerning neither the *good life*, nor supererogatory actions, but rather a contractualists conception of proportioning the interests of oneself and others, cf. BAUMARD NICOLAS/SHESKIN MARK, Partner Choice and the Evolution of a Contractualist Morality, in: J. Decety, T. Wheatley (eds.), The Moral Brain: A Multidisciplinary Perspective, 2015, p. 37. They states that people are not bad utilitarians, failing to maximise group welfare, but rather successful contractualists in regards to the notion of allocating welfare in a fair way, cf. BAUMARD/SHESKIN, (Fn. 27), p. 39.

28 Cf. KOHLBERG, Stage and Sequence: The Cognitive-Developmental Approach to Socialisation, in: D. A. Goslin (ed.), Handbook of Socialization Theory and Research, 1969, p. 397.

29 KOHLBERG (Fn. 28), p. 398.

30 Justice can be defined «as the interaction of the individual with its social environment in relation to the reciprocity of rights and responsibilities», KOHLBERG (Fn. 28), p. 398. In its most elementary form, justice concerns one-on-one reciprocity, transitioning (with scaling difficulty) to a familial, collective, and ultimately social level, cf. GARZ (Fn. 10), p. 25. This will be apparent when subsequently confronted with his stage theory.

31 «The more the social stimulation, the faster the rate of moral development», KOHLBERG (Fn. 28), p. 402.

32 *Hard stage* models are to be differentiated from *soft stage* models. The prior have four distinct

these stages through logical, rational reasoning, transitioning from an egocentric preconventional level to a sociocentric postconventional level.[33] A schematic overview of the mechanisms involved can be found in Figure 1 (see Appendix).[34]

Preconventional Level[35]

Stage 1: Comprehension is oriented to immediate punishment and obedience.
Stage 2: Instrumentally purpose-oriented, «tit for tat».

Conventional Level

Stage 3: Reciprocal interpersonal expectations and relationships, sociological communicative role conditions.
Stage 4: Subject-subject relationships replaced by subject-system relationships, conformity with the law and social institutions.

Postconventional Level

Stage 5: Law as a social contract, the individual's role in the subject-system view.
Stage 6: Orientation on universal, moral principles, derived through a thought experiment similar to RAWLS' veil of ignorance[36] (here: «moral musical chairs»[37]).

The progression through the stages is correlational to the subject's age, and it can be noted that most do not make the transition to the postconventional level.[38]

D. Mentalism and the Universal Moral Grammar Theory

A more recent cognitive approach can be found in the mentalist theory.[39] Morality, it claims, is based on a higher set of universal principles generating moral judgement, akin to the higher language faculty manifesting itself as spoken language.[40] Drawing in large parts from the works of JOHN RAWLS and

characteristics: (1) qualitative differences, (2) hierarchical integration, (3) fixed order of development, and (4) clear distinguishability, cf. GARZ (Fn. 10), p. 33 ff. For an alternate, slightly modified account of the six stages (structured in four *schemas*), cf. GIBBS (Fn. 26), p. 41, 60 ff., 75 ff.

33 Cf. LEVINE C./KOHLBERG L./HEWER A., The Current Formulation of Kohlberg's Theory and a Response to Critics, vol. 8, no. 2, p. 1 f.

34 KOHLBERG regarded the underlying mechanisms to be of cognitive nature, claiming that »the moral force in personality is cognitive. Affective forces are involved in moral decisions, but affect is neither moral nor immoral. When the affective arousal is channeled into moral directions, it is moral; when it is not so channeled, it is not. The moral channeling mechanisms themselves are cognitive«, KOHLBERG LAWRENCE, From Is to Ought: How to Commit the Naturalistic Fallacy and Get Away with It in the Study of Moral Development, in: T. Mischel (ed.), Cognitive Development and Epistemology, 1971, p. 230 f.

35 The following account draws from GARZ (Fn. 10), p. 39–46.

36 Cf. RAWLS JOHN, A Theory of Justice, rev. ed., 1999 p. 118 ff.

37 Cf. KOHLBERG LAWRENCE, The Claim to Moral Adequacy of a Highest Stage of Moral Judgement, in: The Journal of Philosophy, vol. 70, no. 18, 1973, p. 644.

38 Cf. GARZ (Fn. 10), p. 46. The stage model faces several challenges: The observable phenomenon of stage regression, a problematic Stage 4.5, the lack of empirical evidence regarding Stage 6, and a hypothesised cosmic Stage 7. For a deeper dive, consult GIBBS (Fn. 26), Chapter 3, 4.

39 The mentalist theory is hereby regarded a *cognitive* theory due to its hints towards the cognitive nature of the moral faculty, cf. MAHLMANN (Fn. 13), p. 580; DELTON ANDREW W./KRASNOW MAX M., Adaptationist Approaches to Moral Psychology, in: J. Decety, T. Wheatley (eds.), The Moral Brain: A Multidisciplinary Perspective, 2015, p. 21. However, it is conceptually distinct from other *cognitive* approaches, e.g. KOHLBERG'S.

40 Cf. MAHLMANN (Fn. 13), p. 579 f.; MAHLMANN (Fn. 9), p. 123. The distinction between moral *competence* and *performance* is essential, cf. CHOMSKY (Fn. 25); MAHLMANN (Fn. 14), p. 76; ATRAN SCOTT/GINGES JEREMY, Devoted Actors and the Moral Foundations of Intractable Intergroup Conflict, in: J. Decety, T. Wheatley (eds.), The Moral Brain: A Multidisciplinary Perspective, 2015, p. 78; MIKHAIL JOHN, Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgement, 2011, p. 51 ff. This is derived from RAWLS' linguistic analogy, cf. RAWLS, A Theory of Justice, p. 40 ff.

NOAM CHOMSKY, it sets the foundation for a *universal moral grammar*.[41]

The universal moral grammar theory has two distinct components. The *moral grammar* component is based on the notion that every natural language contains words to express certain non-reduceable deontic operators, which constitute the distinct framework of the human mind and behavior.[42] This moral grammar is deemed *universal* because some of its core elements, such as the moral *ought,* or foundational principles of justice and altruism, are regarded innate to the human mind.[43] This is derived from the *poverty of stimulus* argument.[44]

III. Emotive Account

A. Moral Dumbfounding

JONATHAN HAIDT provides a diametrical approach to cognitive models of morality. Moral judgement, he claims, is a relativistic phenomenon constructed by emotion and society.[45] He instigates the point by highlighting the phenomenon of *moral dumbfounding*:[46]

Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that, was it OK for them to make love?[47]

Most participants immediately proclaim the act to be immoral, yet have a hard time justifying this belief, ultimately resorting to statements such as «I don't know, I can't explain, I just know it is wrong».[48] He criticises KOHLBERG's exclusive use of reason-heavy dilemmas such as the *Heinz dilemma*, categorizing them as *moral reasoning tasks* which shroud the complete picture – namely one with «greater prominence to moral emotions and [...] moral intuitions».[49]

B. The Social Intuitionist Model

Moral judgement,[50] HAIDT claims, stems from quick and automatic moral intuitions

41 Cf. MAHLMANN (Fn. 13), p. 579 f.; MIKHAIL JOHN, Universal Moral Grammar: Theory, Evidence, and the Future, in: Trends in Cognitive Sciences, vol. 11, no. 4, 2007, p. 143; DELTON, KRASNOW, *Adaptationist Approaches*, p. 20.

42 Cf. MIHKAIL (Fn. 41), p. 144 ff.; MAHLMANN (Fn. 9), p. 122. This innate moral faculty would enable moral judgements such as *impermissible*, *permissible,* and *obligatory* to be generated, cf. DELTON/KRASNOW (Fn. 39), p. 21.

43 For a closer look at the mentioned ontogenesis of morality, cf. MAHLMANN (Fn. 13), p. 605. The term *innate* is to refer to a system that is «largely pre-determined by the inherent structure of the mind, but whose ontogenetic development must be triggered and shaped by appropriate experience and can be impeded by unusually hostile learning environments», MIKHAIL (Fn. 41), p. 144.

44 The poverty of stimulus argument states that if a certain cognitive ability cannot be generated by outside stimuli, «at least some of the cognitive structures underlying this ability must be inborn», MAHLMANN (Fn. 9), p. 133. Developmental psychological studies revealed infants and young children capable of making the moral/conventional distinction, providing support towards the inertness of the foundational principles of morality, cf. DELTON/KRASNOW (Fn. 39), p. 21.

45 Cf. GIBBS (Fn. 26), p. 1; TASSY/LE COZ/WICKER (Fn. 14), p. 680.

46 This term originates from BJÖRKLUND F./HAIDT J./SCOTT M., Moral Dumbfounding: When Intuition Finds no Reason, Unpublished Manuscript, 2000. They considers moral dumbfoundedness to be «a state in which "seeing-that" conflicts with "reasoning-why"», BJÖRKLUND/HAIDT/SCOTT (Fn. 46), p. 11.

47 HAIDT (Fn. 3), p. 814.

48 HAIDT (Fn. 3), p. 814; BJÖRKLUND/HAIDT/SCOTT (Fn. 46), p. 10 f.; PRÉTÔT LAURENT/BROSNAN SARAH, The Evolution of Morality: A Comparative Approach, in: J. Decety, T. Wheatley (eds.), The Moral Brain: A Multidisciplinary Perspective, 2015, p. 14.

49 BJÖRKLUND/HAIDT/SCOTT (Fn. 46), p. 11.

50 He broadly defines moral judgement as «evaluations (good versus bad) of the actions or character of a person that are made with respect to a

(Link 1 in Figure 2, see Appendix),[51] whereas cognitive moral reasoning[52] serves the subsidiary role of post-hoc justification (Link 2).[53] Additionally, he proposes that moral judgement should be regarded as an interpersonal process (Link 3, Link 4), putting emphasis on the social and cultural influences, referring to moral emotions as «those emotions that are linked to the interests or welfare either of society as a whole or at least of persons other than the judge or agent».[54] These claims are purely descriptive, giving concise account on how moral judgement develops, not how it *ought* to develop.[55] For a complete overview of this model and its intricacies, consult *Figure 2*.

HAIDT backs his social intuitionist model with following four arguments:

(1) Empirical research has shown that «the perception of a person or an event leads instantly and automatically to a moral judgement without any conscious reflection or reasoning», operating through the intuitive judgement link (Link 1).[56]

(2) Only under very specific conditions is it possible to partake in reasoned judgement and private reflection (Link 5 and 6); in a conventional setting, post-hoc reasoning is deployed (Link 2).[57]

(3) Post-hoc reasoning (Link 2) is not a memory of the cognitive processes underlying behavior, but rather a post-hoc justification based on *a priori moral theories*, as is evident from people not being able to identify the reasoning underlying their decision.[58]

(4) Illustrated on the basis of psychopaths, moral emotions – unlike moral reasoning – seem to supply the necessary *ought*, dictating moral behavior.[59]

IV. Dualistic Account

A. Approach

Around the year 2001, efforts were made to locate a *moral module* in the brain, but they have since failed.[60] In what soon turned out

set of virtues held to be obligatory by a culture or subculture», HAIDT (Fn. 3), p. 817.

51 Moral intuition is regarded as «the sudden appearance in consciousness of a moral judgment, including an affective valence (good-bad, like-dislike), without any conscious awareness of having gone through steps of search, weighing evidence, or inferring a conclusion», HAIDT (Fn. 3), p. 818.

52 Being the adversary to moral intuition, moral reasoning is defined as a «conscious mental activity that consists of transforming given information about people in order to reach a moral judgment», HAIDT (Fn. 3), p. 818. Only very rarely does moral reasoning lead to reasoned moral judgement (Link 5); it is rather the result of one's intuition-based judgement (Link 2). GREENE advocates a narrower definition of moral reasoning, suggesting it to be a cognitive process and creating a direct link between subject A's and subject B's capacity to reason, as opposed to the indirect link of A influencing B's intuition (Link 3), cf. PAXTON JOSEPH M./ GREENE JOSHUA D., Moral Reasoning: Hints and Allegations, in: Topics in Cognitive Science, vol. 2, no. 3, 2010, p. 517, 525.

53 Cf. HAIDT (Fn. 3), p. 815, 818; GREENE JOSHUA D., The Secret Joke of Kant's Soul, in: W. Sinnot-Armstrong (ed.), The Neuroscience of Morality: Emotion, Brain Disorders, and Development, vol. 3, 2008, p. 36.

54 HAIDT JONATHAN, The Moral Emotions, in: R. J. Davidson, K. R. Scherer, H. H. Goldsmith (eds.), Handbook of Affective Sciences, 2003, p. 853.

55 Cf. HAIDT (Fn. 3), p. 815.

56 Cf. HAIDT (Fn. 3), p. 818.

57 Cf. HAIDT (Fn. 3), p. 820 ff. HAIDT concedes that the mind has dual-processing properties in (1), yet argues that reasoned judgement rarely takes place in everyday scenarios, cf. GREENE (Fn. 53), p. 36. In comparison, the DPT regards moral reasoning as a «ubiquitous feature of moral common sense», cf. PAXTON/GREENE (Fn. 52), p. 513.

58 HAIDT defines *a priori moral theories* as a «pool of culturally supplied norms for evaluating and criticising the behavior of others», HAIDT (Fn. 3), p. 822. According to HAIDT, reasoning is not only post-hoc, but also biased through «coherence motives» and «relatedness motives», HAIDT (Fn. 3), p. 821 f. See also PRÉTÔT/BROSNAN (Fn. 48), p. 14 f.

59 Cf. HAIDT (Fn. 3), p. 823 ff. It can be observed that moral action leads to internal gratification as a reward for altruistic behavior. Upon closer inspection, a paradoxical property of morality becomes clear: moral actions result in the satisfaction of the moral agent only if said agent intends the well-being of others, as opposed to his own, cf. MAHLMANN (Fn. 14), p. 293 f.

60 It is now widely believed that the neural mechanisms behind moral cognition also participate in

to be a hallmark study, GREENE et al. sought out to find the neural mechanisms behind moral judgement.[61]

Subjects were presented a wide array of moral and non-moral dilemmas and observed through the lens of an fMRI scan. The moral dilemmas were divided into *personal* and *impersonal* categories,[62] involving variations of the so-called *Trolley Problem*.[63] The *standard trolley* scenario is as follows:

You are at the wheel of a runaway trolley quickly approaching a fork in the tracks. On the tracks extending to the left is a group of five railway workmen. On the tracks extending to the right is a single railway workman. If you do nothing the trolley will proceed to the left, causing the deaths of the five workmen. The only way to avoid the deaths of these workmen is to hit a switch on your dashboard that will cause the trolley to proceed to the right, causing the death of the single workman. Is it appropriate for you to hit the switch in order to avoid the deaths of the five workmen?[64]

In the *footbridge* variant,[65] there is no longer a fork and the switch is replaced by a large man on a footbridge over the tracks. It is then questioned whether it is appropriate to push the stranger off the bridge onto the tracks below, stopping the trolley with his large body and saving the five workers.

B. Results

In the *standard trolley* scenario, most participants regarded it appropriate to flip the switch.[66] In the *footbridge* scenario, however, most people chose not to push the large man.[67] The fMRI scans revealed the *footbridge* scenario to recruit subject's emotions to a higher degree, while judgements concerning the *standard trolley* scenario were found to be more closely resembling those of non-moral dilemmas.[68] Subject's reaction times showed emotionally incongruent responses in the moral-personal condition (e.g., when participants responded «appropriate» to the *footbridge* scenario) to take longer than emotionally congruent responses. The other two conditions (moral-impersonal and non-moral) exhibited a trend in the opposite direction.[69]

A new set of experiments further distinguished between easy moral-personal dilemmas and difficult moral-personal dilemmas based on subject's reaction times.[70] Fo-

other capacities, cf. GREENE JOSHUA D./HAIDT JONATHAN, How (and Where) does Moral Judgement Work?, in: Trends in Cognitive Sciences, vol. 6, no. 12, 2002, p. 523.

61 GREENE J. D. et al., An fMRI Investigation of Emotional Engagement in Moral Judgement, in: Science, vol. 293, no. 5537, 2001.

62 Personal dilemmas involve actions that « (a) could reasonably be expected to lead to serious bodily harm, (b) to a particular person or a member or members of a particular group of people, (c) where this harm is not the result of deflecting an existing threat onto a different party», GREENE et al. (Fn. 61), p. 2107 f. en. 9.

63 Its modern form was first introduced by FOOT PHILIPPA, The Problem of Abortion and the Doctrine of the Double Effect, in: Oxford Review, no. 5, 1967, p. 6. The problem had been formulated before, a minor difference being the use of a train instead of a trolley, cf. WELZEL HANS, Zum Notstandsproblem, in: Zeitschrift für die gesamte Strafrechtswissenschaft, vol. 63, no. 1, 1951, p. 51.

64 Cf. Supplemental Data to GREENE et al. (Fn. 61). There has been criticism as to why this variant – suggesting the subject to be at the wheel of the trolley – was regarded as impersonal and not personal, cf. PARDO MICHAEL S./PATTERSON DENNIS, Minds, Brains, and the Law: The Conceptual Foundations of Law and Neuroscience, 2014, p. 59 f.

65 This variant was first introduced by THOMSON JUDITH JARVIS, Killing, Letting Die, and the Trolley Problem, in: The Monist, vol. 59, no. 2, 1976, p. 207 f.

66 GREENE et al. (Fn. 61), p. 2105.

67 GREENE et al. (Fn. 61), p. 2105.

68 GREENE et al. (Fn. 61), p. 2107. Decisions in the moral-personal condition were coupled with higher activation of brain areas associated with emotion (Brodmann Area 9, 10, 31, 39) and significantly lower activation in areas associated with working memory (BA 7, 40, 46) compared to decisions in the moral-impersonal and non-moral conditions. For a compilation of alternate findings, cf. SCHLEIM (Fn. 18), p. 189 ff.

69 GREENE et al. (Fn. 61), p. 2107.

70 GREENE J. D. et al., The Neural Bases of Cognitive Conflict and Control in Moral Judgement, in: Neuron, vol. 44, no. 2, 2008, p. 392. The aim was to »test the hypothesis that different patterns of neural activity in response to the same class of moral dilemma are correlated with dif-

cusing solely on the difficult moral-personal dilemmas, they found increased dorsolateral prefrontal cortex activity (associated with cognitive control) when participant's chose «appropriate» as opposed to «inappropriate»; a choice they consider to be in line with utilitarian decision making.[71]

C. The Dual-Process Theory

The findings were compiled into a theory coined the Dual-Process Theory: moral cognition, so the Dual-Process Theory, stems from both intuitive emotional responses and more controlled cognitive responses.[72] In special situations – as is the case with moral dilemmas such as the trolley problem – they play competing roles.[73] This process has been visualised in Figure 4 (see Appendix).

GREENE compares the human brain to an SLR camera, which can operate in two complementary modes (automatic and manual).[74] It is supposed to exemplify an «elegant solution to the ubiquitous design problem, namely, the trade-off between efficiency and flexibility».[75] The automatic mode is subconscious, guided by reflexes and intuition, serving us well in our day-to-day life; the underlying processes are emotional.[76] The manual mode is dedicated to general-purpose reasoning; it operates on a conscious level and allows us to recognize and adhere to certain rules.[77]

The line between conceptual claims and descriptive conclusions is blurred in regards to the Dual-Process Theory, with GREENE stating to intend only the latter.[78] The normative content that follows will be evaluated below.

V. Foundation of the Criminal Law

A. Purpose

The criminal law can best be characterised as an act-guiding system, deployed to maintain certain societal standards.[79] It rests on several pillars and underlying maxims, some of which have explicitly been codified in Section 1.02 of the Model Penal Code.[80]

ferences in moral decision-making behavior«, GREENE et al. (Fn. 70), p. 390.

71 GREENE et al. (Fn. 70), p. 391 f. «Judgements that maximize aggregate welfare» and «accepting a personal moral violation in favor of a greater good» were named as utilitarian decisions, GREENE et al. (Fn. 70), p. 390, 392. Contradictory results regarding the reaction times and DLPFC activity have since been found, cf. MOORE A. B./CLARK B. A./KANE M. J., Who Shalt not Kill? Individual Differences in Working Memory Capacity, Executive Control, and Moral Judgement, in: Psychological Science, vol. 19, no. 6, 2008, p. 556.

72 Cf. GREENE et al. (Fn. 61), p. 2107. GREENE clarifies the *dual* in Dual-Process to concern the type of processing, as cognitive outputs typically mirror their underlying processes, cf. GREENE JOSHUA D., Beyond Point-and-Shoot Morality: Why Cognitive Neuroscience Matters for Ethics, in: Ethics, vol. 124, no. 4, 2014, p. 697.

73 Cf. GREENE JOSHUA D., The Cognitive Neuroscience of Moral Judgement and Decision Making, in: J. Decety, T. Wheatley (eds.), The Moral Brain: A Multidisciplinary Perspective, 2015, p. 203 f. An alternate explanation was provided by MIKHAIL in form of an act-tree, differentiating between means and side-effects akin to the Doctrine of Double Effect (*Figure 3*, see Appendix). This explanation is consistent with at least twelve variations of the trolley problem, cf. MIKHAIL (Fn. 41), p. 146 ff.; MIKHAIL (Fn. 40), p. 115 f., 118 ff.

74 This is used as an analogy for the Dual-Process Theory, cf. GREENE (Fn. 72), p. 696.

75 GREENE (Fn. 72), p. 696. MIKHAIL acknowledges this problem, but deems moral judgement to be on a level of complexity where simple deontological and consequentialist principles do not provide sufficient explanation, cf. MIKHAIL (Fn. 40), p. 103.

76 GREENE (Fn. 72), p. 696.

77 GREENE (Fn. 72), p. 696 f. «In short, manual mode thinking is the kind of thinking that we think of as "thinking"», Greene (Fn. 72), p. 697.

78 See GREENE et al. (Fn. 61), p. 2107; GREENE et al. (Fn. 70), p. 398. He regards it «an empirical hypothesis concerning a general trend rather than a conceptual claim», GREENE J. D., MORELLI S. A., LOWENBERG K., NYSTROM L. E., COHEN J. D., Cognitive Load Selectively Interferes with Utilitarian Moral Judgement, in: Cognition, vol. 107, no. 3, 2008, p. 1145, fn. 1.

79 Cf. WILSON WILLIAM, *Criminal Law*, 5. edt., 2014, p. 4.

80 The guiding principle being the prevention of harm to individual or public interests, as evident from Section 1.02(1)(a) MPC. Other functions

The resemblance to morality becomes apparent at first glance. Both the criminal law and morality constitute a system of values designated to influence behavior.[81] However, there are discrepancies regarding their *tools of guidance*. While the criminal law imposes a wide array of external, government-induced sanctions for failing to follow conduct,[82] morality avails to inward-facing, psychological means of punishment.[83]

Justifying the necessary measures to fulfil the criminal law's act-guiding purpose has resulted in the development of punishment theories, most prominently those of consequentialism and retributivism.[84]

B. Theories of Punishment

The consequentialist theory justifies punishment by its beneficiary future consequences, namely deterrence and incapacitation; punishment is regarded as a *prima facie* wrong, only justified in light of excluding greater evil.[85] Built on utilitarian soil, the forthbringing of greater social benefit than social harm provides sole grounds for justification.[86]

On retributivist grounds, punishment is regarded as the justified reaction to wrongdoings of the actor, regardless of its future benefits.[87] It has intrinsic worth and is to be proportioned to the desert of the actor.[88]

There are also mixed accounts, incorporating elements of both consequentialism and retributivism. Most notably, HART distinguishes between three justificatory issues, namely (1) the aim, (2) subject, and (3) intensity of punishment.[89] While the aim ought be utilitarian (1), he deems it impermissible to deliberately punish the innocent (2) or excessively punish the guilty (3).[90]

Impacting our understanding of these two philosophical strands would lead to a fundamental change in the criminal law doctrine, and it is the science of moral cognition – namely the Dual-Process Theory – that claims to be able to do just that.[91]

C. The Normative Impact of Science

The issue at hand is as follows: the scientific method can reveal the way things *are*, yet this observation states not what *ought* to be.[92] Assuming the Dual-Process Theory to be a true account of the operations underlying moral judgement, it is not yet clear how this

include rehabilitation, retribution and restoration.

81 Cf. MORSE STEPHEN J., Criminal Law and Common Sense: An Essay on the Perils and Promise of Neuroscience, in: Marquette Law Review, vol. 99, 2015, p. 50, fn. 32.

82 Cf. WILSON (Fn. 79), p. 6 ff.

83 Cf. MAHLMANN (Fn. 14), p. 293 f.

84 Cf. WILSON (Fn. 79), p. 53 ff.; PARDO/PATTERSON (Fn. 64), p. 179 f.

85 «All punishment in itself is evil. [...] if it ought at all to be admitted, it ought only to be admitted in as far as it promises to exclude some greater evil», BENTHAM JEREMY, An Introduction to the Principles of Morals and Legislation, repr., 1907, Chapter XIII §1 II; GREENE JOSHUA D./COHEN JONATHAN, For the Law, Neuroscience Changes Nothing and Everything, in: Philosophical Transactions of the Royal Society B, vol. 359, no. 1451, 2004, p. 1776.

86 Cf. WILSON (Fn. 79), p. 58; PARDO/PATTERSON (Fn. 64), p. 183 f.

87 GREENE (Fn. 53), p. 50 f.; PARDO/PATTERSON (Fn. 64), p. 184 f.; DUFF ANTHONY/HOSKINS ZACHARY, Legal Punishment, in: E. N. Zalta (ed.), The Stanford Encyclopedia of Philosophy, 2014.

88 Cf. WILSON (Fn. 79), p. 54. One ought to punish the criminal because he deserves it, see KANT IMMANUEL, The Metaphysics of Ethics, 1796, J. W. Semple (trans.), H. Calderwood (ed.), 1886, p. 104. For a proposal towards a wholly desert-based criminal code, see ALEXANDER L./FERZAN K. K./MORSE S. J., Crime and Culpability: A Theory of Criminal Law, 2009.

89 Cf. HART H. L. A., Punishment and Responsibility, 1968, p. 1–27.

90 Cf. HART (Fn. 89), p. 1–27.

91 There are many dissenting opinions, most notably that of STEPHEN MORSE, who believes the law ought not be phased by recent neuroscientific findings, cf. MORSE STEPHEN J., Neuroethics: Neurolaw, in: University of Pennsylvania Law School, Public Law Research Paper no. 17/9, 2017, p. 22. However, he concurs that if science *were* able to fundamentally impact those foundational notions, the criminal law ought take notice, cf. MORSE (Fn. 91), p. 45.

92 Recall HUME'S dictum, quoted below the preface.

leads to a normative conclusion, and if it does, to which.

When it comes to changing the criminal law doctrine, (at the very least) three arguments have to be supplied. First, empirical findings (in this case the fMRI results) need to be linked to a mantling theory (Dual-Process Theory). A further argument needs to connect said theory to a normative claim, e.g. deontology being faulty. The criminal law can be reached when deontology is linked to retributivism. Alternatively, the empirical evidence could point towards free will being an illusion, and the criminal law could be reached through subsequent linking of free will to retributivism.

In both examples, the is-ought barrier is seemingly crossed. How can the conclusion »the criminal law ought not punish in retributivist terms« be reached from mere brain scans? The answer is hidden in the modalities of said *ought*. It is used as a shorter form of »should do XY because it is the *right* thing to do«.[93] There is a certain belief component attached, which in turn has to be based on something. For example, the normative claim «you ought not punch others» means «you should not punch others because it is the *wrong* thing to do».[94] This can be reduced to the belief that punching others is wrong because it causes *harm* to them.[95] If science (or any other form of empirical proof) were to show that getting punched by others is beneficial to one's physical health and mental well-being (say, by awakening dormant healing abilities and causing the releasing dopamine and serotonin), that normative claim could be considered largely debunked.

Empirical observations that verify or falsify foundational notions of normative claims thus have an indirect influence on the claim itself.[96]

VI. Direct Approach to Rejecting Retributivism

A. Linking Emotion and Deontology

GREENE observes that when decisions are made with *cognitive*[97] regions of the brain, they result in characteristically consequentialist judgement; when it comes to judgement that is in line with characteristically deontic principles – namely being justified in terms of rights and duties – the areas of the brain responsible for emotive responses are active.[98]

Going back to the camera example, he considers this dichotomy between the automatic and manual mode to highlight that both have their respective strengths and weaknesses and should be used accordingly.[99] The automatic mode is fast and efficient, yet requires prior trial-and-error experiences to shape it, lest its well-functioning would be

93 This «right» is not to be understood in terms of justice, as it is debatable whether acting just is an imperative property of morality. Rather, «right» is synonymous to «correct in regards to one's conviction» (as it is also debatable whether there are any *universally* true/right/correct notions of guidance).

94 «Wrong» does not have to equal «bad» if the normative claim above was postulated as a universal law. A punch can be «good», e.g. as self-defence, yet it would still be regarded as «wrong».

95 The moral relativist might object and ask why reducing harm is desirable in the first place.

96 This thought is also shared by SINGER, who suggests that scientific advances »do not themselves imply any normative conclusions, but [they] undermine some conceptions of doing ethics which themselves have normative conclusions«, SINGER PETER, Ethics and Intuitions, in: Journal of Ethics, vol. 9, no. 3/4, 2005, p. 349.

97 Within the boundaries of the DPT, the term *cognitive* is not to be understood synonymous to *information processing*, as emotions also involve information processing. Rather, it is used as an antonym to *emotion*, in a sense that it does not succumb to automatic behavioral responses, involving an unbiased reasoning process, cf. GREENE (Fn. 53), p. 40.

98 Cf. GREENE (Fn. 73), p. 203; GREENE (Fn. 53), p. 37 ff. His terminological use of consequentialism and deontology is not congruent with standard philosophical usage, which is why he refers to it as «characteristically consequentialist» and «characteristically deontological». A statement is «characteristically consequentialist» when it is justified by utilitarian cost-benefit reasoning and harder to justify in deontological terms, and vice versa (with «characteristically deontological» judgement being justified in terms of rights and duties), cf. GREENE (Fn. 72), p. 699 f.

99 Cf. GREENE (Fn. 72), p. 714.

akin to a «cognitive miracle».[100] Thus, while judging based on deontic yardsticks is most practicable in everyday life, one should rely on manual mode (implying cognitive reasoning) when it comes to dealing with unfamiliar[101] moral problems; he terms this the *No Cognitive Miracles Principle*.[102] This tension, according to his *Central Tension Principle*, stems from the inherent disparity in cognitive design between efficiency and flexibility.[103]

Instead of localizing a first principle and deriving an answer from there, deontology is akin to deriving the first principle from the intuitively right answer.[104] These claims are backed by a large body of independent research.[105]

100 These trial-and-error experiences can result from genetic transmission, cultural transmission or personal experience, as «these are the only mechanisms known to endow human automatic cognitive processes with the information they need to function well», GREENE (Fn. 72), p. 714.

101 Unfamiliarity can be the result of recent cultural development or moral disagreement. In the latter case, the «conflicting intuitions» causing the disagreement should be dropped for the sake of using manual mode, so GREENE (Fn. 72), p. 716 f., 725. Why? Because our automatic mode is not equipped to tackle unfamiliar moral problems and make good intuitive judgement, regardless of what is meant by *good*, cf. GREENE JOSHUA D., Reply to Driver and Darwall, in: S. M. Liao (ed.), Moral Brains: The Neuroscience of Morality, 2016, p. 174. Good automatic judgement without prior shaping experience is a paradox – or in GREENE's words – a cognitive miracle.

102 GREENE (Fn. 72), p. 715. Deontology is regarded as unfit as it displays an emotional affinity towards *mere* personal force and *mere* spatial proximity. The insignificance of personal force is highlighted by pointing out that it should not make a difference morally whether the man in the footbridge scenario was instrumentalised through physical force (pushing him off) or a switch (opening a trapdoor below), cf. GREENE (Fn. 72), p. 713. In regards to spatial proximity, GREENE deems it paradox that we regard it deeply wrong to abandon a bleeding stranger on the side of the road (even if it would result in us having to replace our leather car seats), yet we do not feel an obligation to save the lives of countless people in impoverished parts of the world through a donation of lesser or equivalent value than the to-be-replaced seats, cf. GREENE JOSHUA D., From Neural "Is" to Moral "Ought": What are the Moral Implications of Neuroscientific Moral Psychology?, in: Nature Reviews, vol. 4, 2003, p. 848. For a further read on the paradoxy of the situation, cf. SINGER PETER, Famine, Affluence, and Morality, in: Philosophy & Public Affairs, vol. 1, no. 3, 1972. A similar example is provided by PETER UNGER, who compares refusal to donate to (proven) charitable organisations with allowing a trolley to kill a child rather than diverting it, destroying one's precious vintage Bugatti in the process, cf. UNGER PETER, Living High and Letting Die: Our Illusion of Innocence, 1996, p. 135 ff.

103 Deontic judgement is regarded as a remnant of our emotion-driven primal past, operating the quick, automatic responses for the up-close scenarios our ancestors faced, hereby being nothing more than the rationalisation of our intuitive emotional behavior, the «"cognitive" expression of our deepest moral emotions», GREENE (Fn. 72), p. 699; GREENE et al. (Fn. 70), p. 398; GREENE (Fn. 53), p. 62 f.

104 GREENE (Fn. 72), p. 725. This is illustrated in a striking experiment: ordinary folk and professional philosophers are presented with cases similar to *footbridge* (harm as means) and *switch* (harm as side effect) in mixed order. When participants (ordinary folk and philosophers alike) were presented *footbridge* first, they were biased towards answering *switch* to be impermissible as well. When presented with *switch* first, the Doctrine of Double Effect was evoked 50 % more frequently in the philosophers, providing a congruent answer to *footbridge* (conforming to the differentiation between means and side effect). Further questioning revealed that the philosophers adjusted the theory underlying their decision to be consistent with their choice in the scenarios; thus, the philosophers that deemed *footbridge* and *switch* impermissible (in that order) committed a mistake somewhere, as they failed to invoke the Doctrine of Double Effect. What are the implications? The Doctrine of Double Effect, GREENE claims, is not the underlying principle where judgement is derived from, but simply the codification of our intuitive judgement. For a detailed account on the difficulty of *biting this metal bullet*, cf. GREENE (Fn. 72), p. 719 ff.

105 GREENE (Fn. 72), p. 700 ff. The weight lies not on the individual studies, but their entirety, as «each piece, taken in isolation, is open to alternative interpretations», GREENE (Fn. 72), p. 706. For further studies he cites, cf. GREENE (Fn. 53), p. 41 ff.; GREENE (Fn. 73), p. 204 ff.

B. Linking Deontology and Retributivism

GREENE knows that deontology and retributivism are two conceptually distinct moral theories, yet he regards them virtually indistinguishable within the domain of punishment.[106] Two arguments are supplied:

(1) Consequentialism is seen as the antagonist to both deontology in the field of morality and retributivism in the field of punishment theories. Theories which are opposed to the consequentialist account in the latter field are founded on notions of retributivism to some degree.[107] Thus, strengthening the consequentialist position leads non-consequentialists – and with that retributivists – to shaky ground.

(2) Traditional proponents of deontology tend to endorse punishment on retributivist grounds.[108]

A wide array of evidence is provided for this claim.[109]

C. The Deontic Core of the Mental Gizmo Thesis

MAHLMANN termed the DPT the *mental gizmo thesis* and regarded it a fundamental attack against human rights.[110] Several arguments were brought forth, most notably that it suffers from an internal contradiction. The utilitarian judgement, which is founded on the principle of utility, is considered to be a *slow thinking* process. The principle of utility prescribes everyone's happiness to count equally, thus proposing two conditions: (1) everyone to be regarded equal, and (2) equal persons deserving equal treatment. The second notion is a deontic – rather than consequentialist – principle, and it lies in the heart of utilitarianism. Thus, according to MAHLMANN, the DPT refutes itself by claiming deontic judgement to be *fast thinking*, when it is this very same judgement that lies in the heart of utilitarianism, which GREENE regards as *slow thinking*.[111]

GREENE has explicitly stated that deontological judgement can also result from *slow thinking* (cognitive) processes, yet one does not typically reach a *characteristically deontological* conclusion this way, but rather from intuitive emotional responses (which manifests

---

106 Cf. GREENE (Fn. 53), p. 78 en. 6.

107 Cf. GREENE (Fn. 53), p. 75.

108 Cf. GREENE (Fn. 53), p. 75. This notion can best be observed in regards to IMMANUEL KANT.

109 Studies revealed ordinary people to be conceptually inclined with consequentialist principles of punishment (as means of deterrence), yet sway to the retributivist account of desert when confronted with a tangible case, cf. CARLSMITH K. M., DARLEY J. M./ROBINSON P. H., Why Do We Punish? Deterrence and Just Deserts as Motives for Punishment, in: Journal of Personality and Social Psychology, vol. 83, no. 2, 2002, p. 286 f., 289, 292 f. Participants were found to disregard consequentialist means even when directly confronted with them, punishing in proportion to the *emotional outrage* they felt, cf. GREENE (Fn. 53), p. 50–55. Neuroimaging studies of the ultimatum game confirmed these results, with participants choosing to punish as an ends itself – in spite of no deterring impact – followed by increased activation of the anterior insula and caudate nucleus (both brain regions associated with emotions), cf. RILLING J. K. et al., The Neural Correlates of Theory of Mind within Interpersonal Interactions, in: NeuroImage, vol. 22, no. 4, 2004, p. 1697, 1700 ff. A cross-cultural study on the moral condemnation of harmless actions revealed higher socio-economic status (=SES) and age – thus more developed *cognitive* capacities – to lead to consequentialist responses (not condemning harmless actions). The opposite (low SES and young age) lead to characteristically deontological decision making, cf. GREENE (Fn. 53), p. 55 ff. Moral judgement could better be predicted by offensiveness than harmfulness, cf. HAIDT J./KOLLER S. H./DIAS M. G., Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?, in: Jounal of Personality and Social Psychology, vol. 65, no. 4, 1993, p. 624 f. While the majority of scenarios tested for do not provide a direct link between the condemnation of harmless actions and deontology, the results are identical in the *broken promise* case, which GREENE labels as a case of «downtown deontology», GREENE (Fn. 53), p. 57.
The studies are thus primarily a confirmation of the *cognitive* nature of consequentialism and should be apprehended cautiously for means of linking emotional, punitive punishment to deontology.

110 Cf. MAHLMANN (Fn. 9), p. 111.

111 Cf. MAHLMANN (Fn. 9), p. 116 f.; MAHLMANN (Fn. 14), p. 274 f.

in the significantly lower reaction times).[112] *Characteristically consequentialist* judgement, on the other hand, can never be reached by automation; it is always the result of a slower weighing process recruiting distinct areas of the brain (dorsolateral prefrontal cortex).[113] MAHLMANN deems this self-refuting, regarding the very core of utilitarianism, namely the principle of utility, to be derived from a deontic notion.

The latter must not hold true. The innate, natural expression of human psychology – namely that of justice – satisfies the impartiality requirement of morality without having to fall back to deontic notions.[114]

However, is this »innate, natural expression of human psychology« used to justify consequentialism not based on the very same emotional intuitions GREENE criticises deontology for? Not necessarily. GREENE never claimed consequentialist judgement to be void of emotion, quite the contrary: GREENE sympathises with HUME's allegation that all moral judgement has an affective basis.[115] GREENE differentiates between the «alarm-like», emotional urge of deontological judgement and the consequentialist weighing process, which, while it too is subject to emotion, takes these into account as relevant factors.[116] This reminds of SIDGWICK's solution, who differentiated between perceptual, dogmatic, and philosophical intuition, avoiding the predicament by assigning the intuition underlying consequentialism to the more sophisticated, latter kind.[117]

112 GREENE (Fn. 53), p. 65; for his account on the reaction times, cf. GREENE et al. (Fn. 61), p. 2107; GREENE et al. (Fn. 70), p. 390; PAXTON, GREENE (Fn. 52), p. 521 f.

113 GREENE (Fn. 53), p. 64.

114 Cf. MILL JOHN STUART, Utilitarianism, 1863, in: John Stuart Mill, Utilitarianism, Batoche Books (ed.), 2001, p. 41 ff. The usual objection is dismissing MILL'S account as fallacious, yet there is good reason to regard it as deductively valid, cf. MILLGRAM ELIJAH, Mill's Proof of the Principle of Utility, in: Ethics, vol. 110, no. 2, 2000.

115 Cf. GREENE et al. (Fn. 70), p. 397; GREENE (Fn. 53), p. 64.

116 Cf. GREENE (Fn. 53), p. 64.

117 See SIDGWICK HENRY, The Methods of Ethics, 1874, J. Bennett (ed.), 2015, Chapter 8; GREENE, Ethics, p. 724.

D. Deontology as Heuristics

GREENE makes it clear that the frequent accusations of him being opposed to emotion-based moral judgements are false, as he attributes strengths and weaknesses to both modes (automatic and manual).[118]

This seems to miss the crux of such accusations. One might imagine a computer with two pre-installed programs: Program A is able to create vivid, captivating stories based on the input of a few key words. Program B is able to solve any mathematical question posed, akin to a calculator. It can now be said that the computer has two different programs with their respective strengths and weaknesses, which is what GREENE claims the dual-processing mind to have. If, however, Program B were also capable of conceiving stories of the same calibre as Program A, only taking longer to do so, it would vastly discredit Program A.

It is clear that heuristic-like judgements have a strong practical benefit for everyday life, as we cannot ponder about every miniscule decision. However, assigning efficiency as the only advantage means *settling* for automatic mode not due to its superiority in that field, but real-life practicability: if one had enough time at their disposal, the factually correct answer to any question would result from manual, cognitive thinking. The frequent accusations turn out to be true, as GREENE is discrediting deontology on a conceptual basis by writing it off as heuristics.

E. Interim Conclusion

It seems that both retributivism and deontology share common roots in emotional soil – one that consequentialism is not based on.[119] Judging by the premise that these

118 Cf. GREENE (Fn. 72), p. 714. Automatic mode relying on deontology and manual mode on consequentialism.

119 As elaborated in section VI.C, GREENE is inclined to follow HUME's account that all moral judgement stems from affect; it is the differentiation between »alarm-like« emotions and an emotional weighing process that is crucial in this regard, and what is meant with *emotional soil*.

roots are indeed capable of discrediting deontology, bridging the gap to retributivism requires little extra effort.[120] Nevertheless, it is a daring assumption to make.

VII. Indirect Approach to Rejecting Retributivism

A. Presupposed Picture of the Law

The criminal law regards persons as conscious and rational agents with the capacity to enact control over their own actions (so called *practical reasoners*).[121] This view houses on a folk psychological framework: action is causally explained through certain *mental* – not brain – *states* such as desires, beliefs, intentions, and plans.[122]

When talking about actions and behavior, we consider a certain someone – a *person* pulling the strings – to be in charge. While it is up for debate whether we should narrow this personhood down to humans only, there is mutual agreement on agency being a necessary attribute ascribed to said person.[123] In the criminal law's retributivist sense, only those agents that can be blamed for their actions deserve punishment.[124] Thus, the presupposed picture is that of a *free* agent.

B. The Problem of Free Will

This notion of free will entails the ability to do otherwise, to act as an uncaused causer.[125] It stands in contrast to determinism: the idea that the world in its current state «is completely determined by (i) the laws of physics, and (ii) past states of the world».[126] How do the two relate?

---

120 PARDO/PATTERSON claim that this does not undermine even a subset of retributivist views, as correlating a theory to emotional areas of the brain is a mere observation and does not prove the theory to be incorrect (only showing how things *are*, not how they *ought to be)*. GREENE, they claim, provides no independent criteria that determines consequentialist reasoning to be *more correct* than its deontic counterpart, see PARDO/PATTERSON (Fn. 64), p. 189 f.
This view is a strawman. While GREENE did suggest deontology to be on the wrong path towards «moral truth», he never claimed consequentialism to have discovered such either. Instead, he simply regards consequentialism as the current «best available standard for public decision making», GREENE (Fn. 53), p. 77; also GREENE (Fn. 101), p. 175. Deontological decision making is discredited by its link to moral intuition, thus prone to morally irrelevant influences (e.g., mere spatial proximity).

121 Cf. WILSON (Fn. 79), p. 34; WALDBAUER JACOB R./GAZZANIGA MICHAEL S., The Divergence of Neuroscience and Law, in: Jurimetrics, vol. 41, no. 3, 2001, p. 359. As MORSE states: «Legally responsible agents are therefore persons who have the general capacity to grasp and be guided by good reason in particular legal contexts», MORSE STEPHEN J., Neuroscience and the Future of Personhood and Responsibility, in: University of Pennsylvania Law School, Public Law Research Paper no. 12/26, 2011, p. 117.

122 MORSE STEPHEN J./CATHERINE JOHN D., A Primer on Criminal Law and Neuroscience: A Contribution of the Law and Neuroscience Project, 2013, p. xxxiii. This is reflected in the criteria for criminal culpability: the actus reus, mens rea, and affirmative defences are all dependent on the mental states of the agent, see MORSE STEPHEN J., Lost in Translation? An Essay on Law and Neuroscience, in: M. Freeman (ed.), Law and Neuroscience, vol. 13, 2011, p. 530 f. While biological and sociological variables also influence behavior, folk psychology considers mental states to be the essential causal link, see MORSE (Fn. 121), p. 117.

123 A distinct feature of this personhood being their «status as a morally responsible agent» and a special kind of control exclusive to them, see ESHLEMAN ANDREW, Moral Responsibility, in: E. N. Zalta (ed.), The Stanford Encyclopedia of Philosophy, 2016. This thought can be taken further, questioning whether the status of being a person is exclusive to humans or if it includes or ought to include certain (nonhuman) animals as well. DARWIN made a famous conjuncture regarding this, saying that »any animal whatever, endowed with well-marked social instincts, would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well developed, or as nearly developed, as in man«, DARWIN CHARLES, The Descent of Man, and Selection in Relation to Sex, 1871, in: J. T. Bonner, R. M. May (eds.), The Descent of Man, and Selection in Relation to Sex, 1981, p. 71 f.

124 MORSE STEPHEN J., Psychopathy and Criminal Responsibility, in: Neuroethics, vol. 1, no. 3, 2008, p. 208.

125 GREENE/COHEN (Fn. 85), p. 1777; MORSE (Fn. 91), p. 16; MORSE (Fn. 81), p. 54.

126 GREENE/COHEN (Fn. 85), p. 1777. There are different types of determinism, but for the sake

There are three main approaches to this conundrum, namely those of hard determinism, libertarianism, and compatibilism.[127] Hard determinism deems free will and determinism to be incompatible, advocating the latter to be true.[128] Libertarianism, too, is built on incompatibilist grounds, yet it regards determinism to be false.[129] Compatibilism, as the most common view among professional philosophers,[130] presumes the previous two approaches to be conceptually mistaken in regards to free will: rather than being an uncaused causer, «agents must simply have the capacity to determine their actions by reasons and to act in light of those reasons and are not compelled to act in the ordinary meaning of compulsion».[131] The criminal law operates under this compatibilist framework of «practical rationality» even in absence of genuine free will.[132] Causation is thus not a per se excusing condition and has to be linked to compromised rationality; believing otherwise would be what MORSE terms the «fundamental psycholegal error».[133]

C. The Curious Case of Mr. Puppet

GREENE/COHEN maintain the position that the dualist position of the criminal law cannot be upheld.[134] What MORSE regards as the «fundamental psycholegal error», they understand as the gap between folk intuition and the compatibilist view of the law.[135] The example of *Mr. Puppet* aims to highlight this discrepancy:

Mr. Puppet is a hypothetical person that has been genetically scripted by a scientist who sought after designing a human to commit a specific crime. This scientist controlled every variable of Mr. Puppet's life, every single event – from his infancy to his teenage years – with 95 % accuracy. Just as the scientist predicted, Mr. Puppet committed said crime. Can Mr. Puppet be deemed guilty?[136]

1. Approaching from Neuroscience

The Mr. Puppet argument presupposes determinism and targets the law's compatibilist stance. For proof of said determinism, one might point towards the studies of BENJAMIN LIBET, which revealed subject's decisions to be made 350–400 milliseconds before they were consciously aware of their intention to act.[137] LIBET adds that there was a timeframe of 100 milliseconds between conscious intention and performance, during which the subject could assert a «veto» over his decision.[138]

Free will seems largely debunked. After all, how much freedom is really left when human decisions once deemed their own are revealed to be that of their subconscious? Our consciousness would turn out to be a mere byproduct of subconscious brain activ-

of this discussion, the focus is placed on causal determinism.

127 GREENE/COHEN (Fn. 85), p. 1777; MORSE (Fn. 81), p. 45.

128 GREENE/COHEN (Fn. 85), p. 1777.

129 GREENE/COHEN (Fn. 85), p. 1777.

130 Survey results showing that 56% are either compatibilist or tend to compatibilism, see BOURGET DAVID/CHALMERS DAVID J., What do Philosophers Believe?, in: Philosophical Studies, vol. 170, no. 3, 2014, p. 476, 490.

131 MORSE (Fn. 81), p. 48.

132 PARDO, PATTERSON (Fn. 64), p. 199. Criminal responsibility is thus not dependent on the notion of free will and fully compatible with determinism, cf. MORSE (Fn. 122), p. 533.

133 MORSE STEPHEN J., Neuroscience, Free Will, and Criminal Responsibility, in: University of Pennsylvania Law School, Public Law Research Paper no. 15/35, 2015, p. 262.

134 The kind of mind-body dualism concerned with here is that of the mind as a non-physical mental entity that has an impact on the physical realm, see GREENE/COHEN (Fn. 85), p. 1784.

135 GREENE/COHEN (Fn. 85), 1777. They argue that the law ought to reflect the «moral intuitions and commitments of society», GREENE/COHEN (Fn. 85), p. 1778.

136 GREENE/COHEN (Fn. 85), p. 1780.

137 LIBET BENJAMIN, Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action, in: The Behavioral and Brain Sciences, vol. 8, no. 4, 1985, p. 529.

138 LIBET (Fn. 137), p. 529, 537 f. This veto ought not be understood as a kind of *free will*, but rather as a *free-won't*, see HAGGARD PATRICK, Neuroethics of Free Will, in: J. Illes, B. J. Sahakian (eds.), The Oxford Handbook of Neuroethics, 2011, p. 221. A critical analysis on the scientific validity and conclusions can be found in MORSE (Fn. 122), p. 551.

ity on which the agent has no influence – *he* did not make the decision, it was his brain.[139]

This notion has to be rejected. A readiness potential in the brain is far from a decision,[140] and subconscious brain activity preceding conscious one is not proof of determinism, but rather how one would expect the brain to operate.[141] There are also several methodological concerns. A 400 millisecond delay between conscious and subconscious processes for pointing a finger at a clock under strict laboratory settings does not translate into everyday action capacity involving decisions magnitudes more complex. There are scenarios in which humans have to consciously choose to act in a span shorter than 400 milliseconds – the swing/no swing decision of a professional baseball player being an example.[142]

This does not rule out that one day, neuroscience might be able grant full insight into the «mind's clockwork» and reveal our thoughts to be nothing more than red neurons firing against blue neurons.[143] Current neuroscience, however, is still far from that.

2. Approaching from Physics

Determinism – on a conceptual level – seems to be built on false notions of physics. The idea of a fully causal universe is based on Newtonian physics, yet ever since the discovery of quantum mechanics, we know this account to be incomplete. Assuming quantum mechanics to cause truly random events, the world in its current state would be the result of (1) the laws of physics, (2) the past state of the world, and (3) random quantum mechanical events.[144] This randomness seems to be incompatible with determinism.

There are two ways to tackle this argument. First, the premise of quantum mechanics as a truly random force of nature might be mistaken. The accounted randomness of quantum-mechanical events is based on a form of the Copenhagen interpretation,[145] and there is an ongoing discourse in the field as to whether that holds true.[146] For the sake of the argument, we can assume quantum mechanics to be truly random. What follows? Causal determinism would lose its footing, as future events could not be deduced from past world states anymore. This, however, is neither proof of the reality of free will, nor the retributivist precondition of *desert*. Human action – in the sense of an uncaused causer – is no more free in a fully random universe than a fully determined one.

D. The Curious Case of Mr. Puppet Continued

In regards to GREENE/COHEN's argument, we ought to presume determinism to hold at

139 By the time the agent wanted to do something, that decision had «already been made by lower-level brain mechanisms», DENNO DEBORAH W., Crime and Consciousness: Science and Involuntary Acts, in: Minnesota Law Review, vol. 87, no. 2, 2003, p. 327.

140 Cf. MORSE (Fn. 122), p. 550.

141 Cf. FREEMAN MICHAEL, Introduction: Law and the Brain, in: M. Freeman (ed.), Law and Neuroscience, vol. 13, 2011, p. 6.

142 In the scenario of a 90 miles per hour pitch from less than 60 feet away, the batter has less than half a second to make a conscious decision, cf. PARDO, PATTERSON (Fn. 64), p. 129, fn. 31.

143 GREENE/COHEN state the example of a futuristic scanner that might be able to track every neuron and neuronal connection in the brain, analyse the data and visualise the human decision-making process (*red neurons vs blue neurons*), cf. GREENE/COHEN (Fn. 85), p. 1781.

144 Cf. GREENE/COHEN (Fn. 85), p. 1780.

145 This interpretation states that there is no definitive position of a particle prior to its observation; once observed, it leads to a wave function collapse, and only then does it have a concrete state beyond a mere wave function probability, cf. BOHM DAVID, A Suggested Interpretation of the Quantum Theory in Terms of «hidden» Variables. I, in: Physical Review, vol. 85, no. 2, 1952, p. 167 f.

146 There are alternate interpretations of quantum mechanics which do not hold footing in true randomness, advocating a deterministic account instead, for example the De Broglie-Bohm Theorem, see BOHM (Fn. 145), p. 169 ff. Another possible solution – and part of a more recent debate – would be the many-worlds interpretation, which assumes a corresponding universe for each possible state of the wave function to exist, see EVERETT HUGH, The Theory of the Universal Wave Function, 1956, p. 63 ff.

least partially true, as the following question would not arise if definitive proof of free will were present: Can Mr. Puppet be deemed guilty? After all, the law regards him as rational as any other member of society, his actions a reflection of his desires and beliefs.[147] Yet the intuitive reaction is no, Mr. Puppet ought not be blamed, at least not in a retributivist sense.[148] He cannot be regarded *guilty*, as he was not *at fault*; his actions were carried out by him, yet not *his own*. Blaming *him* for what *he did* seems wholly out of place – he was simply a victim of «neuronal circumstances».[149] It is not intended for Mr. Puppet to be exempt from all punishment and free to roam the streets, as consequentialist principles of deterrence and incapacitation would still apply.[150] Only the retributivist component of *genuine* moral blame would be alleviated.

Criticising the law for diverging from the folk's incompatibilist intuition seems hypocritical coming from GREENE, as he now bases an argument off the same moral intuitions he deemed unreliable and biased in regards to deontology.[151] Even if humans were intuitively incompatibilist, one would commit the original naturalistic fallacy to assume this statement to be of any direct normative value.[152] MORSE deems it a form of «neuroarrogance» to expect fundamental notions of human behavior and the law to change based on a science that has not been able to provide a solution to the mind-body problem.[153]

This leads to an important question: Why does the burden of proof lie on the incompatibilist theories? The law's compatibilist account has provided no genuine proof either – the construction of «practical reason»[154] seems to be a placeholder until the mind-body problem is resolved. Would it not make more sense to stray from retributivist notions until humans have been *proven* to be free agents? By asserting blame in state of such uncertainty, the maxim of *in dubio pro reo* is violated – a glaring error when considering that the criminal law's act-guiding purpose could be fulfilled on wholly consequentialist grounds.

VIII. Conclusive Remarks

It is not without reason that the discourse on moral cognition is ongoing in such a fierce manner. After all, the stakes could not be any higher.[155] When GREENE first pre-

147 GREENE/COHEN (Fn. 85), p. 1780.

148 Among the general population, an incompatibilist stance was taken when determinism was presumed true in an alternative universe; when the same question was posed in regards to our universe, common folk shifted towards assigning moral responsibility and blameworthiness, cf. ROSKIES ADINA L./NICHOLS SHAUN, Bringing Moral Responsibility Down to Earth, in: The Journal of Philosophy, vol. 105, no. 7, 2008, p. 376, 381. This hints towards the difficulty of letting go of certain retributivist notions which seem to be driven by a strong affective value.

149 GREENE/COHEN (Fn. 85), p. 1781. «The Astonishing Hypothesis is that "You", your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of the vast assembly of nerve cells and their associated molecules. [...] You're nothing but a pack of neurons», CRICK FRANCIS, The Astonishing Hypothesis: The Scientific Search for the Soul, 1995, p. 3.

150 GREENE/COHEN (Fn. 85), p. 1783.

151 They were framed as heuristics and deemed unreliable in «unfamiliar» situations (No Cognitive Miracles Principle), see section VI.A. The free will debate is as «unfamiliar» as it gets.

152 After all, the determinism debate has been ongoing in philosophy since millennia; it will not be decided by a rather simple observation, cf. MORSE (Fn. 91), p. 16.

153 Cf. MORSE (Fn. 81), p. 67; MORSE (Fn. 122), p. 546 f. As WITTGENSTEIN famously asked: «Wenn "ich meinen Arm hebe", hebt sich mein Arm. Und das Problem entsteht: was ist das, was übrigbleibt, wenn ich von der Tatsache, daß ich meinen Arm hebe, die abziehe, daß mein Arm sich hebt?», WITTGENSTEIN LUDWIG, Philosophische Untersuchungen, 1945, para. 621.

154 This conception of *practical rationality* is dependent on the existence of mental states; it succumbs to the radical determinist notion of eliminative materialism, where all agency ceases to exist, cf. MORSE (Fn. 81), p. 67. While neuroscience did not provide definitive proof towards this materialist account, doubts on dualist theories were raised.

155 «[...] if commonsense intentional psychology really were to collapse, that would be, beyond comparison, the greatest intellectual catastrophe in the history of our species; if we're that wrong

sented his DPT seventeen years ago, he was subject to a plethora of critique. To this day, his essays are being cited upwards of two-hundred times a year and remain integral to the discourse of moral cognition.[156] The potential impact is immense, and if true, would extend to the deepest corners of our lives, the criminal law being a rather proximate one.

However, one should not overplay the findings. Not only is the DPT based on neuroscientific data which is highly susceptible to methodological errors,[157] but it also stands in competition to several other theories – some more sophisticated than itself. The dual-process theory resembles a further argument – albeit a strong one – in the ongoing debate on human thought and morality; a debate that has remained inconclusive since millennia. The battle of HUME and KANT, once exclusive to the domain of the mind, has now been taken to the laboratories.

The criminal law is founded on the same notions that are currently at stake. As science advances, these might change, but for now, lawmakers ought not worry.

about the mind, then that's the wrongest we've ever been about anything. [...] We'll be in deep, deep trouble if we have to give it up.» FODOR JERRY A., Psychosemantics: The Problem of Meaning in the Philosophy of Mind, 1987, p. xii.

156 SCHLEIM (Fn. 18), p. 195.

157 «Voodoo correlations», statistical smoothing, circular analysis, and reverse interference, to name some of the current issues with neuro data, cf. MAHLMANN (Fn. 9), p. 118 f.

IX. Appendix

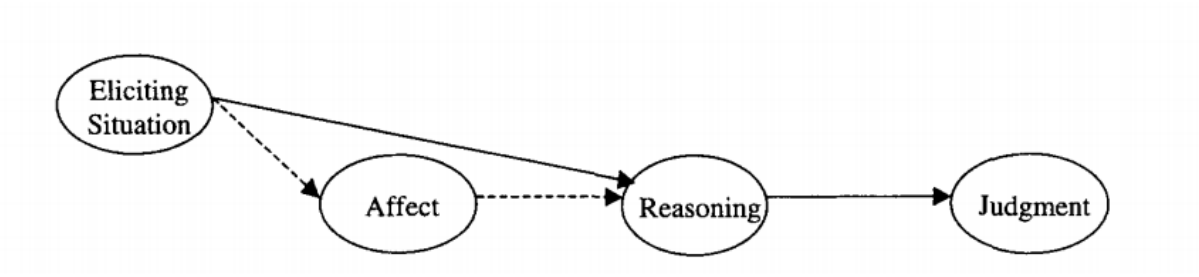

Figure 1, the cognitive reasoning process behind moral judgement; retrieved from HAIDT (Fn. 3), p. 815.

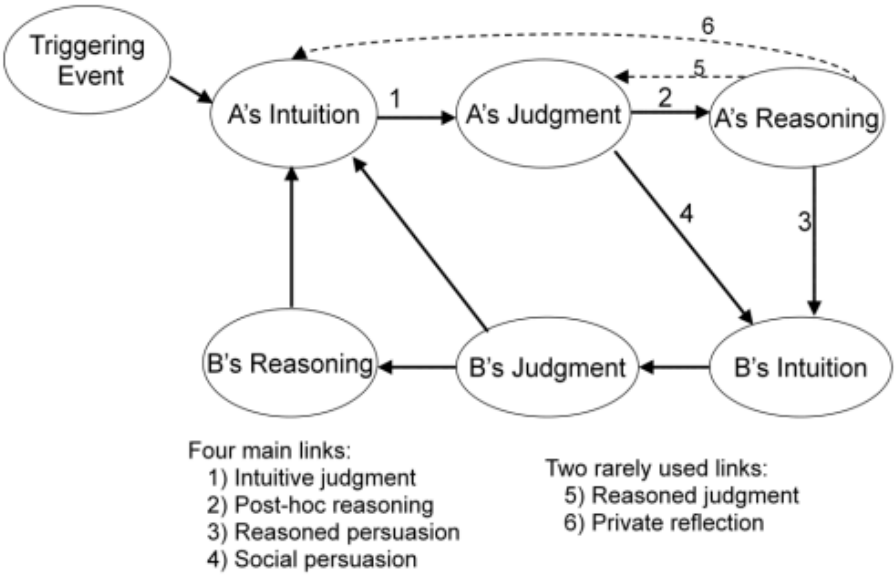

Figure 2, the links of the social intuitionist model; retrieved from HAIDT JONATHAN, Figures for «The Righteous Mind: Why Good People are Divided by Politics and Religion», 2012, Chapter 2 Figure 2.4 (accessed July 27th, 2019).

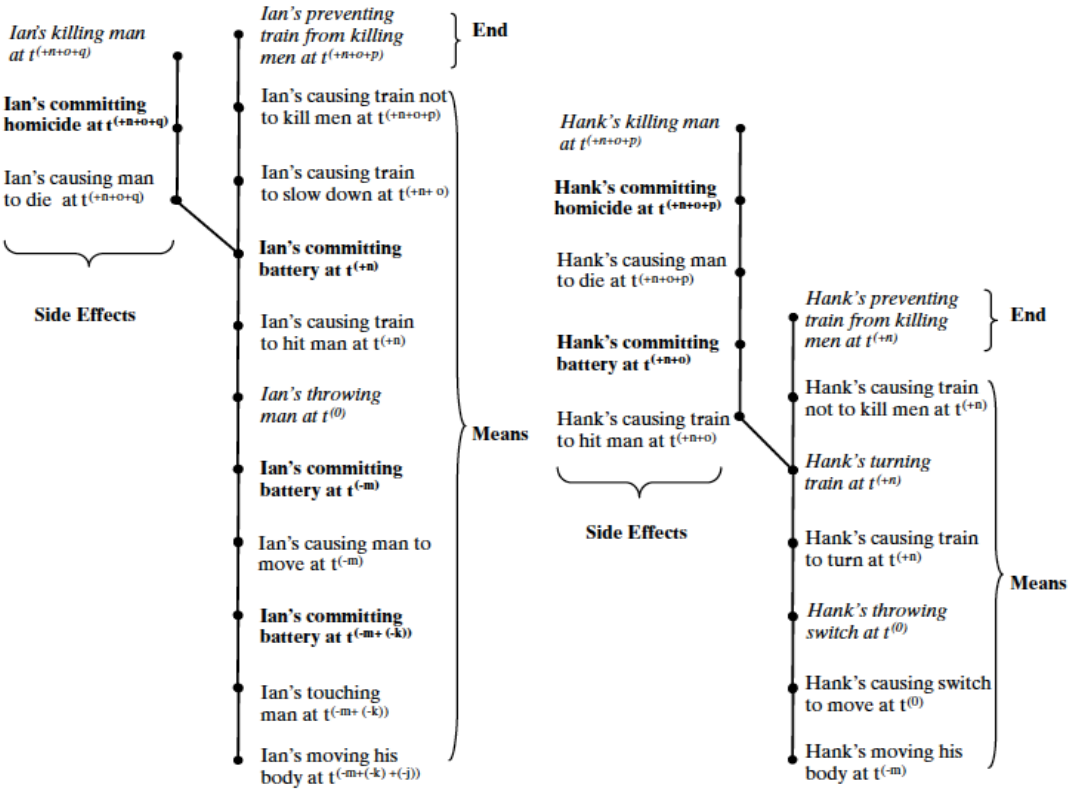

Figure 3, the act-tree approach in regards to *footbridge* (left) and *bystander* (right); retrieved from MIKHAIL (Fn. 50), p. 119.

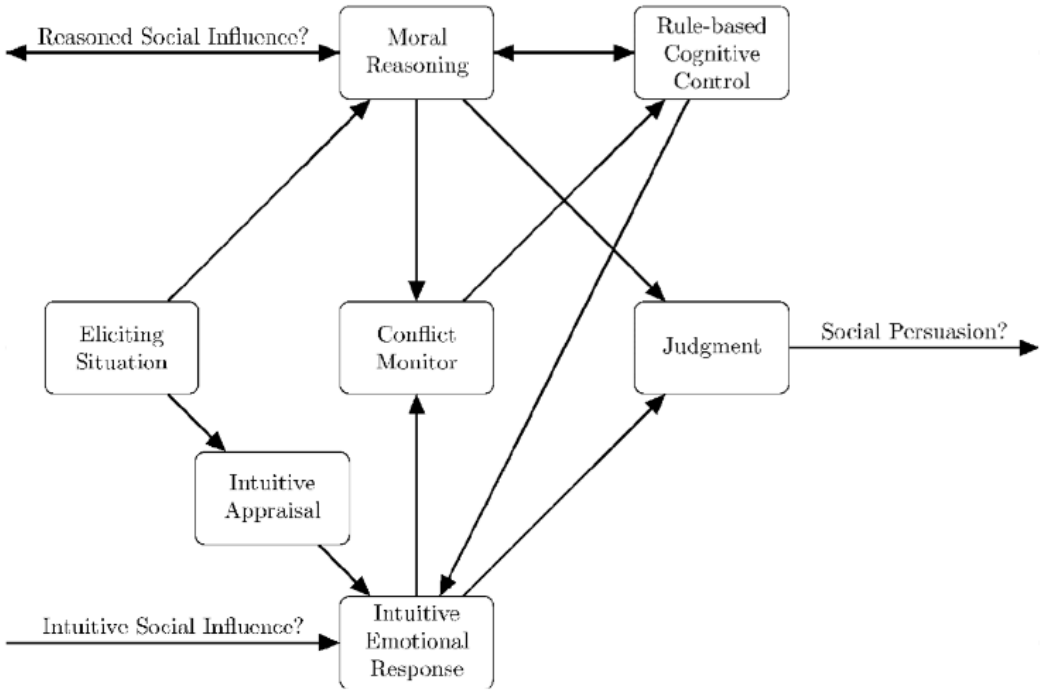

Figure 4, an overview of the Dual-Process Theory; retrieved from PAXTON/GREENE (Fn. 52), p. 514.