TOWARDS A DUAL-PROCESS THEORY OF CRIMINALIZATION

WORK IN PROGESS

Christoph Winter

Harvard University & Humboldt-University of Berlin

Abstract

For centuries, legal philosophers have been trying to find answers to the question: What conduct should be criminalized? While continental European and German scholars in particular tend to focus on the protection of *legal goods*, theorists from common law jurisdictions traditionally favor the *harm principle*. There are numerous differences between and within these and other approaches, yet, their methodology remains the same. Legal philosophers draw heavily on their own intuitions. However, research in behavioral economics, moral psychology and experimental philosophy suggests that some intuitions may be more reliable than others. Hence, this paper is a first attempt to question the prevailing intuitive approach to theorizing criminal law. I will outline how knowledge about the functioning of the human brain can provide valuable information for theories of criminalization focusing specifically on the *dual-process theory of learning and decision-making*. If the empirical findings from the behavioral sciences hold up, it ultimately provides an argument in favor of a *dual-process theory of criminalization* differentiating between traditional and modern criminal laws. From this perspective, human intuitions may still be reliable regarding traditional criminal laws such as murder and assault. Yet, when examining the legitimacy of modern criminal laws such as environmental and animal protection laws or those concerning the application of new technologies such as lethal autonomous weapons systems (LAWS) or gene-editing tools, a consequentialist approach may be the preferred option.

For centuries, legal philosophers have been trying to find answers to the question of what conduct should be criminalized. While continental European and German scholars in particular tend to focus on the protection of *legal goods* ("Rechtsgut"),[1] theorists from common law jurisdictions traditionally favor the *harm principle*.[2] There are numerous differences between and within these and other approaches, such as the theory of the protection of the *rights of others*[3] or the perspectives offered by *legal moralism*[4] and the

---

[1] Proponents of the theory of legal goods include, among others, Claus Roxin, *Der gesetzgebungskritische Rechtsgutsbegriff auf dem Prüfstand*, 160 GOLTDAMMER'S ARCHIV FÜR STRAFRECHT 433-453 (2013); Bernd Schünemann, *Das Rechtsgüterschutzprinzip als Fluchtpunkt der verfassungsrechtlichen Grenzen der Straftatbestände und ihrer Interpretation*, *in* DIE RECHTSGUTSTHEORIE 133-154 (Roland Hefendehl et al. eds., 2003); Winfried Hassemer & Ulfrid Neumann, *Vor § 1 n. 62*, *in* NOMOS KOMMENTAR ZUM STGB, VOL. 1 (Urs Kindhäuser et al. eds., 4th ed., 2013); for a critical analysis of this approach, *see* CHRISTOPH WINTER, METAMORALISCHE KRIMINALISIERUNG 18-74 (forthcoming, 2019).

[2] For defenses of the harm principle, *see*, e.g., JOEL FEINBERG, HARM TO OTHERS (1984); JOSEPH RAZ, THE MORALITY OF FREEDOM ch. 15 (1986); ANDREW ASHWORTH, PRINCIPLES OF CRIMINAL LAW ch. 2 (5th ed., 2006); A. P. SIMESTER & ANDREAS VON HIRSCH, CRIMES, HARMS, AND WRONGS: ON THE PRINCIPLES OF CRIMINALISATION (2011); for many useful clarifications and systematizations of different versions of the harm principle, *see* James Edwards, *Harm Principles*, 20 LEGAL THEORY 253-285 (2014).

[3] TATJANA HÖRNLE, GROB ANSTÖßIGES VERHALTEN. STRAFNORMEN ZUM SCHUTZ VON MORAL, GEFÜHLEN UND TABUS 65–71 (2005); Tatjana Hörnle, *'Rights of Others' in Criminalisation Theory*, *in* LIBERAL CRIMINAL LAW THEORY, ESSAYS FOR ANDREAS VON HIRSCH 169-186 (A. P. Simester, Antje du Bois-Pedain & Ulfrid Neumann eds., 2014).

[4] Proponents of legal moralism include, *inter alia*, MICHAEL S. MOORE, PLACING BLAME (1997) and R.A. DUFF, ANSWERING FOR CRIME: RESPONSIBILITY AND LIABILITY IN THE CRIMINAL LAW 84–89 (2009); R.A. Duff, *Towards a Modest Legal Moralism*, 8 CRIMINAL LAW AND PHILOSOPHY 217-235 (2014); R.A. DUFF, THE REALM OF CRIMINAL LAW (2018) 52-101; Frank Meyer, *Towards a Modest Legal Moralism:*

republican theory.[5] Yet, their methodology remains the same. Philosophers and legal theorists alike draw heavily on their own intuitions. This does not come as a surprise given that it was all an enlightened *Homo sapiens* could rely on in the past. However, recent research from the cognitive sciences including behavioral economics, moral psychology and experimental philosophy suggests that some intuitions may be more reliable than others.

This paper therefore sets out to question the prevailing intuitive approach to theorizing criminal law and offers a first attempt to include scientific findings into the philosophical framework of legal philosophers analyzing the legitimacy of criminal law. More specifically, I will draw on the so-called *dual-process theory of learning and decision-making* which has seen a recent rise in interest since it was popularized in Daniel Kahneman's "Thinking Fast and Slow"[6] and applied to moral decision-making in Joshua Greene's "Moral Tribes".[7] Against this backdrop, the article ultimately introduces the *dual-process theory of criminalization* distinguishing between traditional and modern criminal laws when examining their legitimacy. Accordingly, human intuitions may still

---

Concept, Open Questions, and Potential Extension, 8 CRIMINAL LAW AND PHILOSOPHY 237-244 (2014); for definitory issues of the notion of "legal moralism" itself, *see* Thomas Søbirk Petersen, *What is Legal Moralism?* 12 NORTHERN EUROPEAN JOURNAL OF PHILOSOPHY 80-88 (2011).

[5] JOHN BRAITHWAITE & PHILIP PETTIT, NOT JUST DESERTS: A REPUBLICAN THEORY OF CRIMINAL JUSTICE (1990); Philip Pettit, *Criminalization in Republican Theory*, *in* CRIMINALIZATION: THE POLITICAL MORALITY OF THE CRIMINAL LAW 132-150 (R.A. Duff et al. eds., 2014).

[6] DANIEL KAHNEMAN, THINKING FAST AND SLOW (2012).

[7] JOSHUA GREENE, MORAL TRIBES: EMOTION, REASON, AND THE GAP BETWEEN US AND THEM (2013).

offer a reliable guide regarding age-old criminal laws such as murder and assault. Yet, a consequentialist approach may be preferred pertaining to modern criminal laws such as environmental and animal protection laws or those concerning the application of new technologies such as lethal autonomous weapons systems (LAWS) or gene-editing tools. In this vein, the analysis will also introduce a new behaviorally informed approach to understand the much-debated distinction between *malum in se* and *malum prohibitum*.

To show this, I will first review the importance of intuitions within theories of criminalization (I). I will then introduce the *behavioral approach to law and philosophy* and outline its importance for legal philosophy more broadly and theories of criminalization in particular (II). Based on this, I will examine potential normative implications of the behavioral approach to criminalization and introduce the *dual-process theory of criminalization* (III). I will close by pointing out future research directions and making some remarks on interdisciplinary legal studies (IV).

I. THE INTUITIVE APPROACH TO CRIMINALIZATION

The primary focus of the first section is the examination of whether the principles and methods proposed by criminal law theorists are able to mitigate intuitive criminalization. I will start with the relation between criminal law theory, political philosophy and the role of intuition (1). Thereafter, I will consider further (intuitive) remedies suggested to limit criminalization (2), namely the emphasis on mitigating overcriminalization within the discourse of criminalization itself, the so-called "critical function" of criminal law theory

(a) as well as the tendency to introduce additional principles and qualifying requirements (b).

1) Political Philosophy, Criminalization & Intuition

Although the question what conduct *may*, *should* or *must* be criminalized sounds like *the* most central question in normative criminal law philosophy, it was not until recently that legal philosophers' attention partially shifted from punishment to criminalization theories.[8] The most agreeable and welcoming development during this steep rise of interest may be the call for embedding criminalization theories into the arena of political philosophy.[9] Hörnle, for example, notes "that the very first stage in a theory of criminalization should not yet deal with the moral wrongfulness of conduct but with issues of political philosophy."[10] In the words of Duff, Farmer, Marshall, Renzo and Tadros from their joint introduction to a recent handbook on criminalization theories:

---

[8] To many, this is „surprising". *See*, e.g. Tatjana Hörnle, *Theories of Criminalization*, *in* THE OXFORD HANDBOOK OF CRIMINAL LAW 679 (Markus D. Dubber & Tatjana Hörnle eds., 2014); interestingly, there is no systematic work analyzing the nexus between theories of punishment and criminalization. For short discussions of the relation between the two, *see ibid*., at 686; James Edwards, *Criminalization without Punishment*, 23 LEGAL THEORY 69-95 (2017); Paul McGorrery, *The Philosophy of Criminalisation: A Review of Duff et al.'s Criminalisation Series*, 12 CRIMINAL LAW AND PHILOSOPHY 196-197 (2018).

[9] *See*, e.g., A.P. Simester & Andreas von Hirsch, *On the Legitimate Objectives of Criminalisation*, 10 CRIMINAL LAW AND PHILOSOPHY 370 (2016).

[10] Tatjana Hörnle, *Theories of Criminalization*, 10 CRIMINAL LAW AND PHILOSOPHY 304 (2016)

> "The criminal law […] is a political institution, part of the state, and must address us as citizens - members of the polity whose law it is. A theory of criminalization must therefore include or depend on a political theory of state and society: it must be a theory of the role that criminal law should play within a particular kind of polity."[11]

The lack of interaction between criminal law theory and political philosophy which is particularly prevalent within the German-speaking debate arguably led to parallel discussions on the legitimacy of state action in different fields without reaping the fruits of interdisciplinary exchange.[12] For instance, even widely influential political philosophies such as Rawls' theory of justice[13] or Sen's[14] and Nussbaum's[15] capabilities oriented approaches have not entered the discourse despite their enormous potential to benefit the currently dominating liberal approaches to criminalization offered by the harm principle and the theory of the protection of legal goods (*Rechtsgutslehre)*.[16] To some degree, the distinct development of criminalization theories is rather surprising given its roots in 19th century political philosophy. The harm principle can be traced back to the

[11] R.A. Duff et al., *Introduction – Towards a Theory of Criminalization?*, *in* CRIMINALIZATION: THE POLITICAL MORALITY OF THE CRIMINAL LAW 5 (R.A. Duff et al. eds., 2014).

[12] For an analysis of the associated drawbacks, *see* WINTER, *supra* note 1.

[13] JOHN RAWLS, A THEORY OF JUSTICE (1971).

[14] AMARTYA SEN, DEVELOPMENT AS FREEDOM (1999).

[15] MARTHA NUSSBAUM, CREATING CAPABILITIES (2011).

[16] To my knowledge, merely von Hirsch very briefly referred to Rawls' and Sen's political philosophy despite their obvious connection to liberal criminal law theory in Andrew von Hirsch, *Harm and Wrongdoing in Criminalisation Theory*, 8 CRIMINAL LAW AND PHILOSOPHY 249 (2014).

utilitarian philosopher John Stuart Mill[17] while the origins of the *Rechtsgutslehre* are usually identified in the works of P.J.A. Feuerbach[18] and J.M.F. Birnbaum,[19] whose writings were significantly influenced by Kant and Hegel respectively.

Since the introduction of these theories, much work has been spent on optimization and clarification. Pertaining to the harm principle, extensive examinations have been conducted on what constitutes a *harm* and who counts as *others*. Feinberg, for example, defines harm as a "thwarting, setting back or defeating of an interest"[20] while Simester and von Hirsch focus on "a diminution of the kind of things that make one's life go well" and of an "impairment of resources" in the sense of longer term means or capabilities.[21]

---

[17] The now infamous passage from Mill's "On Liberty" states "that the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others." JOHN STUART MILL, ON LIBERTY 14 (1859). Needless to say, there is no need to favor a utilitarian theory in order to embrace the harm principle as many proponents of liberal criminal law theory have done over the past 150 years. While some commentators take Mill to be more concerned with freedom than with utility in "On Liberty", his own statement at *ibid.,* at 136 could hardly be more clear: "It is proper to state that I forego any advantage which could be derived to my argument from the idea of abstract right, as a thing independent of utility. I regard utility as the ultimate appeal on all ethical questions." For examples of modern versions of the harm principle, *see supra* note 2.

[18] PAUL JOHANN ANSELM FEUERBACH, LEHRBUCH DES GEMEINEN IN DEUTSCHLAND GÜLTIGEN PEINLICHEN RECHTS (11th ed. 1832).

[19]Johann Michael Franz Birnbaum, *Über das Erforderniss einer Rechtsverletzung zum Begriffe des Verbrechens, mit besonderer Rücksicht auf den Begriff der Ehrenkränkung,* 15 ARCHIV DES CRIMINALRECHTS 149-194 (1834).

[20] FEINBERG, *supra* note 2, at 33.

[21] SIMESTER & VON HIRSCH, *supra* note 2, at 36-37.

Regarding the discourse on the notion of "legal goods", one can find an extraordinary amount of different approaches within the diverse schools of the theory. Roxin, for instance, defines legal goods as "given conditions or purposes which are necessary for the free development of individuals, the realization of their fundamental rights, or the functioning of the state based on these goals"[22] whereas Hassemer and Neumann prefer to keep it more simple by referring to "human interests".[23] Without question, it would be over-ambitious to even try to give a fair representation of the numerous and detailed variations of the harm principle, the theory of legal goods or other theoretical approaches to criminalization. At this stage, it is sufficient to join the various theorists who pointed out that existing definitions are so widely interpretable that only very minor restrictions are effectively posed on criminalization efforts.[24] Moreover, the fact that criminalization theories are traditionally focused on what conduct *may* be criminalized rather than on what *ought* to be criminalized, further decreases guidance for legislatures. Combined with the observation that theories lack sufficient clarity to impose actual limitations, this approach opens the floodgates for moral intuitions. Because moral intuitions are not

[22] CLAUS ROXIN, STRAFRECHT ALLGEMEINER TEIL I 16 (4th ed., 2006).

[23] Winfried Hassemer & Ulfrid Neumann, *Vor § 1 n. 144*, *in* NOMOS KOMMENTAR ZUM STGB, VOL. 1 (Urs Kindhäuser et al. eds., 4th ed., 2013).

[24] *See*, among others, DUFF, *supra* note 4, at 138; Edwards, *supra* note 2, at 282; DOUGLAS HUSAK, OVERCRIMINALIZATION: THE LIMITS OF THE CRIMINAL LAW 72 (2009); Hamish Stewart, *The Limits of the Harm Principle*, 4 CRIMINAL LAW & PHILOSOPHY 28-29 (2010); Victor Tadros, *Harm, Sovereignty, and Prohibition*, 17 LEGAL THEORY 50 (2011); WINTER, *supra* note 1.

always reliable, as I will outline later in more detail,[25] the question arises how one might prevent (exclusively) *intuitive criminalization*.

Philipp Pettit acknowledges this issue and argues that a solution "requires reliance on an overall theory of the purpose of government and law."[26] Consequently, he views the embedding of criminalization into the field of political philosophy as a necessity to "avoid […] ad hoc intuitions and judgments."[27] I largely agree with Pettit's reasoning,[28] yet, there are four crucial factors for why I would like to take a slightly different approach here and focus primarily on the question of which intuitions are trustworthy rather than starting from first principles or specific political theories.

First, my hope is that some of the findings will be relevant to different commentators in the debate on criminalization which would strengthen the case for the behavioral approach introduced in the second section. Accordingly, section I and II in this article do not aim to argue for or against specific political philosophies or theories of criminalization. Instead, the behaviorally informed approach to criminalization might offer new perspectives on criminalization for different schools of thought. Given the extraordinary amount of normative uncertainty in the political domain, the engagement with the scientific understanding of intuitions may be a particularly worthwhile endeavor. Secondly, even if a theory of criminalization is based on a sound political philosophy,

[25] *See infra* section II.

[26] Pettit, *supra* note 5, at 132.

[27] *Ibid*.

[28] Also in this vein, WINTER, *supra* note 1, at 12-18.

one might still end up with widely interpretable conditions such as the harm principle. From this perspective, further knowledge about intuitions' reliability may be especially valuable. Thirdly, a political philosophy may itself be (partially) based on intuitions and respective rationalizations in which case these intuitions would be passed on all the way to the more specific theories of criminalization. Hence, a political philosophy would not be the solution to mitigate intuitive judgments but rather a further tool to implement and strengthen the intuitive approach to criminalization on a more abstract level.

The final reason starts with the widely accepted recognition of the existence of overcriminalization.[29] Indeed, if we assume that the search for guiding principles and definitions has not been overly successful in terms of reducing criminalization, a different strategy would be to look at the second stage of the current approach to criminalization, namely the forming of intuitions about what conduct may be criminalized. It requires to ask: What does intuitively count as "harm" or a "human interest"? What actions are intuitively considered "wrong"? And crucially, to what degree are those intuitions trustworthy and are there circumstances under which we should be particularly hesitant to accept the intuitively appealing results?

2) (Intuitive) Limitations to Criminalization

[29] It is rare to find a single academic piece on the topic of criminalization without a reference to Husak's seminal work in HUSAK, *supra* note 24; but *see* also Erik Luna, *The Overcriminalization Phenomenon*, 54 AMERICAN UNIVERSITY LAW REVIEW 703-743 (2005); James Chalmers & Fiona Leverick, *Quantifying Criminalization*, *in* CRIMINALIZATION: THE POLITICAL MORALITY OF THE CRIMINAL LAW 54-79 (R.A. Duff et al. eds., 2014).

Before we dive into the roots, strengths and weaknesses of (moral) intuitions, we need to consider current approaches to limit the scope of criminal law theories because these strategies might already significantly limit intuition-based criminalization. The fact that the following analysis will often refer to and examine liberal approaches to criminalization should neither be taken as an indicator that the approach offered in this article is only of interest to such theories nor should one infer from the critical remarks that I consider liberal theories as inferior. The reason for this focus merely results from the fact that liberal approaches tend to be more critical of the ineffective restrictions on criminalization of present theories, and hence, it is more likely to find intuition-limiting factors here. At the same time, this also means that if no effective mitigating principles can be found here, it is less likely to find them elsewhere which strengthens the case for criminalization being largely based on intuitions.

a) The "Critical Function" of "Human Interest"

A noteworthy phenomenon is that liberal theorists are particularly keen on emphasizing the "critical function" of criminalization theories. The tendency to judge a theory of criminalization based on its ability to question existing criminal laws has been particularly influential in the German-speaking discourse,[30] where Engländer even argues

[30] *See*, *inter alia*, Armin Engländer, *Revitalisierung der materiellen Rechtsgutslehre durch das Verfassungsrecht?*, 127 ZEITSCHRIFT FÜR DIE GESAMTE STRAFRECHTSWISSENSCHAFT (ZStW) 616-621 (2015); Wolfgang Frisch, *An den Grenzen des Strafrechts*, *in* BEITRÄGE ZUR RECHTSWISSENSCHAFT: FESTSCHRIFT FÜR WALTER STREE UND JOHANNES WESSELS ZUM 70. GEBURTSTAG 72 (1993); Winfried

that the "critical potency" (as it is often referred to) determines the theory's "performance capabilities".[31] At first glance, this may be understandable from two perspectives. First, when thinking about criminalization from the point of view of overcriminalization, a theory becomes more appealing, as soon as it tackles this fundamental issue. This appeal may be particularly strong in the context of the current status of the US criminal justice system. Secondly, collectivist approaches tend to extend the legitimate applicability of criminal law even further. Often mentioned examples in this regard are the protection of "public peace" or "security" via the criminal law.[32] Notwithstanding the importance of both public peace and security, such terminologies have been proven to be particularly equipped to support presently held moral convictions. Even Feuerbach, often viewed as

---

Hassemer, *Rechtsphilosophie, Rechtswissenschaft, Rechtspolitik – am Beispiel des Strafrechts*, *in* RECHTS- UND SOZIALPHILOSOPHIE IN DEUTSCHLAND HEUTE 130, 139 (Robert Alexy et al. eds., 1991); WINFRIED HASSEMER, THEORIE UND SOZIOLOGIE DES VERBRECHENS. ANSÄTZE ZU EINER PRAXISORIENTIERTEN RECHTSGUTSLEHRE 53-54 (1973); Young-Whan Kim, *Verhaltensdelikte versus Rechtsgutsverletzungen – Zur aktuellen Diskussion um einen materiellen Verbrechensbegriff*, 124 ZSTW 611 (2012); Hans Kudlich, *Die Relevanz der Rechtsgutstheorie im modernen Verfassungsstaat*, 127 ZSTW 642-644 (2015); Antonio Martins, *Der Begriff des Interesses und der demokratische Inhalt der personalen Rechtsgutslehre*, 125 ZSTW 238-239 (2013); SABRINA PFAFFINGER, RECHTSGÜTER UND VERHÄLTNISMÄßIGKEIT IM STRAFRECHT DES GEISTIGEN EIGENTUMS 89 (2015); THOMAS RÖNNAU, WILLENSMÄNGEL BEI DER EINWILLIGUNG IM STRAFRECHT 26 (2001); PETER SINA, DIE DOGMENGESCHICHTE DES STRAFRECHTLICHEN BEGRIFFS „RECHTSGUT" 26-42 (1962); Sabine Swoboda, *Die Lehre vom Rechtsgut und ihre Alternativen*, 122 ZStW 24-32 (2010); Wolfgang Wohlers, *Die Güterschutzlehre Birnbaums und ihre Bedeutung für die heutige Rechtsgutstheorie*, GA 600-606 (2012).

[31] Engländer, *supra* note 30, at 616-621.

[32] *Cf.* Hörnle, *supra* note 8, at 694 who further notes that "arguments pointing to collective interests lead straight into the area of endangerment offenses."

the founder of German liberal criminal law theory, introduced a separate legal category, the so-called *Polizey-Vergehen*, to protect public order and security. Here, he was able to justify along with the *Zeitgeist* of 19th century Bavaria the (sometimes severe) punishment of, among others, infidelity and drunkenness, which a strictly liberal approach arguably might have prevented.[33]

Although I share the general motivation to put effective limitations on the criminal law, there are at least two issues worth mentioning with the above line of reasoning. First, critique is not *intrinsically* valuable as often implicitly assumed by commentators in the German debate. A theory which only allows criminalization of conduct that causes significant harm to panda bears is certainly more critical than its rivals but the theory nevertheless has substantial shortcomings. Furthermore, despite the fact that liberal theories may in fact be viewed as more critical than collectivist approaches, they do not offer legislatures much guidance themselves. Take for instance Hassemer's influential approach to justify prohibitions imposed by the criminal law based on "human interests" only. Since almost anything may be interpreted as a human interest, the approach does not offer much orientation either. Case in point, liberal proponents within the normative theory of competition law argue that fair or well-functioning competition is a human

[33] Hörnle is a bit less critical of Feuerbach's *Polizey-Vergehen* and notes that they do not entail an „invitation for unbridled moralism", Tatjana Hörnle, *P.J.A. Von Feuerbach and His Textbook of the Common Penal Law (1801)*, *in* FOUNDATIONAL TEXTS IN MODERN CRIMINAL LAW 138 (Markus Dubber ed., 2014). In another article, she argues that collective goods such as "public peace" deserve critical scrutiny "because they might serve as a disguise for moralistic lawmaking", Hörnle, *supra* note 8, at 695.

interest.[34] The protected legal good in this case is not the competition *per se*, but the human "interest" or "trust" in fair and well-functioning competition.[35] Similarly, one might argue for the protection of the human interest in public peace and safety.

Needless to say, the goal of my reasoning is not to argue for or against the protection of the interest in fair and well-functioning competition or public peace and safety via the criminal law. The examples solely illustrate that it can make very little difference to the ultimate justification of the criminal law in question whether one refers to a collectivist or liberal theory due to the broad interpretation of legal goods, harms, interests and similar notions.

Nevertheless, the reader might wonder why I had to pick exactly these examples since liberal approaches are surely known to be very hesitant to accept arguments along the line of public security. However, the reason why liberal theories *in practice* do not argue for the protection of a "human interest in public peace" is irrelevant for the current argument that such an approach does not offer the necessary *theoretical* guidance either. More importantly, this observation helps to outline the current intuitive approach to criminalization. Although there is no reason in principle why liberal criminal law theorists would not accept a "human interest in public security" but a "human interest in fair competition", the intuitive approach can arguably explain why this is the case in practice. More precisely, research in moral psychology indicates that there is a good chance that

[34] *See*, e.g., Olaf Hohmann, *StGB § 298 Wettbewerbsbeschränkende Absprachen bei Ausschreibungen*, *in* MÜNCHENER KOMMENTAR ZUM STGB 1-5 (Wolfgang Joecks & Klaus Miebach eds., 3rd ed. 2019).

[35] *Ibid*.

intuitions about specific criminal laws will be rationalized on the more abstract level of "human interest" rather than starting from first principles and go "wherever those principles happen to lead."[36] As Jonathan Haidt puts it: "Intuition comes first, strategic reasoning second."[37] Because "public security" references have been used to justify illiberal criminal laws, liberal criminal law theorists certainly have a strong motive to deny its status on a more abstract level; in this case as a "human interest". To be clear, I agree with the skeptical approach towards this line of justification, but the basic denial of public security as a human interest altogether further exemplifies the definitory issues of "human interests" and the reliance on intuitions about concrete laws rather than the proposed abstract principles.

b) Qualifying Requirements and Additional Principles

Due to the fact that extensive and vague definitions of widely interpretable theories may not be able to limit criminalization, multiple other remedies have been proposed.[38] Such suggestions may be classified into two different types of strategies. One is to add *qualifying requirements*. Prittwitz, for example, argues that it is not sufficient to justify criminal laws with *any* human interest, but rather one ought to refer to interests of humans

[36] Joshua Greene, *Beyond Point-and-shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics*, 124 ETHICS 725 (2014).

[37] JONATHAN HAIDT, THE RIGHTEOUS MIND - WHY GOOD PEOPLE ARE DIVIDED BY POLITICS AND RELIGION 61-83 (2012); more on this in section II.

[38] To my knowledge, only Hassemer and Neumann generally reject the idea of further criteria, *see* Hassemer & Neumann, *supra* note 23, at n. 145.

living "responsible" and "dignified" lives.[39] Others prefer that interests ought to be "important",[40] "sufficiently significant",[41] "rational"[42] or "fundamental"[43] and Feinberg famously distinguished between "welfare" and "ulterior" interests.[44] Yet, it remains unclear whether criteria which are vague in themselves are in fact able to mitigate overall vagueness and prevent vastly intuition-based criminalization.[45]

The second strategy is based on the introduction of *additional principles* which is more common within the debates surrounding the harm principle than regarding the theory of legal goods.[46] Particularly popular in this regard is the principle of "wrongfulness".[47]

---

[39] Cornelius Prittwitz, *Personale Rechtsgutslehre und die „Opfer von morgen"*, *in* „PERSONALE RECHTSGUTSLEHRE" UND „OPFERORIENTIERUNG IM STRAFRECHT" 110 (Ulfrid Neumann & Cornelius Prittwitz eds., 2007); for a critical note on Prittwitz's additional elements, *see* WINTER, *supra* note 1, at 64-68.

[40] HÖRNLE, *supra* note 3, at 74.

[41] Kudlich, *supra* note 30, at 646.

[42] HÖRNLE, Grob anstößiges Verhalten, S. 74.

[43] Kai Ambos, *The Overall Function of International Criminal Law: Striking the Right Balance Between the Rechtsgut and the Harm Principles*, 9 CRIMINAL LAW AND PHILOSOPHY 319 (2015).

[44] FEINBERG, *supra* note 2, at 28–45.

[45] Even some of the advocates themselves remain skeptical, see Prittwitz, *supra* note 39, at 110.

[46] As usual, there are exceptions, *see*, e.g., Gerhard Seher, *Prinzipiengestützte Strafnormlegitimation und der Rechtsgutsbegriff*, *in* DIE RECHTSGUTSTHEORIE, LEGITIMATIONSBASIS DES STRAFRECHTS ODER DOGMATISCHES GLASPERLENSPIEL? 39-56 (Roland Hefendehl, Andrew von Hirsch & Wolfgang Wohlers eds., 2003).

[47] See, *inter alia*, SIMESTER & VON HIRSCH, *supra* note 2; von Hirsch, *supra* note 16, at 245-256; HUSAK, *supra* note 24, at ch. 2; Victor Tadros, *Wrongness and Criminalization*, *in* THE ROUTLEDGE COMPANION

Accordingly, conduct may be criminalized only if it is wrongful. For example, von Hirsch defends his and Simester's two-element approach[48] relying on both *harmful* and *wrongful* conduct with the argument that this "provides reciprocal limiting principles concerning the scope of criminalization."[49] However, as any existing definition of *wrongfulness* leaves an extensive range for (intuitive) interpretation as well, the criterion may not be equipped to function as the ultimate mitigating principle after all. Moreover, one might run the risk of paving the way for ever more principles such as the notorious *offense principle*[50] which, in turn, might rather expand than reduce the territory of criminalization. Although I do not necessarily consider this slippery slope argument as a decisive factor in favor of sticking to one foundational principle, it has some force which should not be neglected.

TO PHILOSOPHY OF LAW 165 (Andrei Marmor ed., 2012); *see* also R.A. DUFF, *supra* note 4, at 83-88 who interprets the wrongfulness criterion in Feinberg's harm principle as a ''side constraint of justice''; a comprehensive list of scholars endorsing the wrongfulness criterion in one way or another and not merely as an additional side-constraint can be found in Andrew Cornford, *Rethinking the Wrongness Constraint on Criminalisation*, 36 LAW AND PHILOSOPHY (2017) 615–649 arguing that "theorists of criminal law almost unanimously endorse this principle" (615); Legal moralist, of course, take the wrongfulness of an action not as a limiting factor to criminalization but rather as the (primary) source of its justification.

[48] SIMESTER & VON HIRSCH, *supra* note 2.

[49] Von Hirsch, *supra* note 16, at 245-256; for an examination of the relationship between the "harms" and the "wrongs" in von Hirsch and Simester's work, *see* John Stanton-Ife, *What is the Harm Principle for?*, 10 Criminal Law and Philosophy 329-353 (2016) arguing that the role of the harm principle ultimately becomes a minor one.

[50] JOEL FEINBERG, OFFENSE TO OTHERS (1985).

3) Preliminary Summary

Two widely recognized issues within the debate on the limits of the criminal law are the immense overcriminalization and the lack of sufficiently guiding normative theories due to vague terminologies. While multiple strategies including the embedding of criminalization theories within the field of political philosophy as well as additional principles and qualifying requirements have been proposed, vagueness and, thus, equal reliance on intuition seems to remain. At the same time, the intuitive approach to criminalization has hardly been questioned from a behavioral perspective. Hence, this will be the focus of section II.

II. A BEHAVIORAL APPROACH TO LAW, PHILOSOPHY & CRIMINALIZATION

Starting with a thought experiment, I will first outline the value of the behavioral approach to law, philosophy and criminalization (1) after which I will make some remarks about related developments in the fields of behavioral law & economics and experimental jurisprudence (2). I will then introduce dual-process theory (3) and sketch out two behaviorally informed normative arguments for theories of criminalization, namely the argument from cognitive biases and the argument from the theory of evolution (4) before I will focus on the normative implications of dual-process theory more specifically in section III.

1) The Value of the Behavioral Approach to Legal Philosophy

Imagine you are stranded on a desert island and there is no easy way to leave or seek for help otherwise. But unlike Tom Hanks in the classic *cast-away* scenario, there are many people on the island with you, who desperately want to leave. Moreover, instead of a volleyball you receive all kinds of materials and tools. The people soon realize that the material can be used to build a gigantic ship. You do not know whether the ship will be able to carry all the people, but a large number of the islanders would surely be saved. You start putting things together, but soon realize that nobody on the island knows how to use some of the more advanced tools which are necessary to build the ship. Without this knowledge, you are only making slow progress by building smaller, canoe-like boats which could carry at least some people but not as many as a much greater ship could. You prefer to engage in a trial-and-error approach to building the gigantic ship anyway for some time when, suddenly, a manual for the more advanced tools washes up on shore. Unfortunately, the salt water did some damage to the manual and only some information regarding the tools' usage is available. Would you have a look at it?

If you decide to use the manual, I would argue that one should also look at the tools humans employ in order to build theories of criminalization. More specifically, a closer examination of the insights provided by the behavioral sciences during the last two decades on the functioning of the human brain may turn out to be beneficial. Just as the manual concerning the advanced tools' usage was not delivered in its best possible shape and was incomplete, there are still many gaps in our knowledge about the decision-making processes of the brain. However, recent advances from a variety of fields including behavioral economics, cognitive psychology and neuroscience may still

provide valuable information which brings us closer to understanding human thought processes. Equipped with this knowledge, we may also be able to build better theories of criminalization. Bearing in mind that, just like the manual which only explains how to use the tools, not how to build the gigantic ship, science will not provide us with ultimate answers on how to develop an ideal theory of criminalization. Nevertheless, it may deliver insights which enable us to understand the cognitive tools we use to improve criminalization theories and, by that, might get us a bit closer to solving the puzzle.

One might question the decision to focus on building the gigantic ship because progress has slowly been made by building the smaller canoes. Maybe, if one just takes the time to build more and more small boats, everyone would ultimately be able to leave the island as well. Indeed, slow but steady (moral) progress has been made in many domains of human life[51] including criminalization theories. It therefore seems plausible to infer that more progress will be made in the future as well. However, the argument I put forward is not suggesting to *exclusively* adopt a behavioral approach to criminalization. I merely aim to show that instead of focusing *all* attention on the present intuitive approach to criminalization, it may be worthwhile for researchers to have a look at the brain's manual provided by the behavioral sciences to see whether some steps along the way can be improved. Even if this will not enable one to build the gigantic ship or *the* correct theory of criminalization, maybe the new tools can improve some of the canoes and theories in safety and size.

[51] STEVEN PINKER, THE BETTER ANGELS OF OUR NATURE (2011).

Notwithstanding that significant human progress has overall been made so far, there were multiple tremendous setbacks in human history causing immense suffering. While some of these setbacks were caused by active human choices to engage in warfare and the like, others were the consequence of human inability to mitigate unanticipated risks. One example for this is the agricultural revolution occurring in multiple places around the world around 10,000 years ago. Although I do not suggest that, all things considered, it was a bad choice for *Homo sapiens* to settle, that choice did have some disastrous effects, such as large-scale spreads of diseases, leading to premature deaths of hundreds of thousands, if not millions, of people.[52] A multi-methodological approach can mitigate unknown risk to some degree as the existence of different tools make it more likely that at least one of them will be able to identify and solve the unknown problem. For example, if we only focus on building small canoes, unanticipated weather changes may be fatal whereas larger boats may still be able to set sail. If one thinks that the present technological revolution, including advances in artificial intelligence, geo-engineering, brain-computer interfaces and many other domains, are likely to involve unknown risks, there is even more reason to favor a multi-methodological approach.

Furthermore, some of the major setbacks have not been caused by human inability to solve newly emerging challenges, but were created by human decisions themselves. One can think of an infinite number of examples throughout time and space of what humans nowadays consider moral atrocities ranging from colonialism to slavery and the traditional burning of widows (*sati*). Even philosophers who carefully thought about these

[52] YUVAL NOAH HARARI, SAPIENS 101-125 (2014).

and other issues came to conclusions which strike the vast majority of people today as clearly wrong. Kant, for instance, argued for racist and sexist positions and justified the killing of extramarital children.[53] Even utilitarian philosophers like John Stuart Mill and Henry Sidgwick who argued for women's rights, animal protection laws and the decriminalization of homosexuality long before their academic rivals, were sometimes mislead by their intuition and, among many others, justified British colonialism.[54] Strikingly, Mill and Bentham did not justify their arguments in favor of women's rights or the decriminalization of homosexual acts by referring to their intuitions. On the contrary, Bentham explicitly stated:

> "I have been tormenting myself for years to find if possible a sufficient ground for treating them [homosexual men] with the severity with which they are treated at this time of day by all European nations [i.e. death]: but upon the principle of utility I can find none."[55]

[53] *Cf.*, *inter alia*, IMMANUEL KANT, THE METAPHYSICS OF MORALS 277-284, 314-320, 336, 425 (Mary Gregor transl., 1996).

[54] *See*, e.g., Duncan Bell, *John Stuart Mill on Colonies*, 38 POLITICAL THEORY 34-64 (2010); Bart Schultz, *Mill and Sidgwick, Imperialism and Racism*, 19 UTILITAS 104–130 (2007); but *note* also JEREMY BENTHAM, EMANCIPATE YOUR COLONIES: ADDRESSED TO THE NATIONAL CONVENTION OF FRANCE AS 1793 (1830).

[55] Jeremy Bentham, *Offences against One's Self: Paederasty (Part 1),* 1785, JOURNAL OF HOMOSEXUALITY 389 (1978); a similar passage regarding women's rights can be found in JOHN STUART MILL, ON THE SUBJECTION OF WOMAN (1869). I owe both of these references to Joshua Greene.

Thus, Bentham relied on his utilitarian principles rather than on his own intuitions, which would have led to very different results. This is not to suggest that some version of utilitarianism is *the* solution regarding criminalization or political philosophy more broadly construed, it merely exemplifies that some of the philosophers who came to moral positions far ahead of their time did so by arguing against their own intuitions rather than relying on them. The existence of countless examples of past intuitions favoring what nowadays intuitions consider as moral atrocities should motivate us to ask why one would think that our present intuitions are reliable.[56] What sort of criminal laws will be considered as morally outrageous by future generations, and under what conditions may one actually rely on (moral) intuitions? These and related questions lie at the core of the behavioral approach.

2) From Behavioral Law & Economics to Behavioral Law & Philosophy

Although the behavioral approach is new to theories of criminalization, it is not an entirely new development within the realm of legal studies. In fact, there are a few subfields in which progress has already been made. Most notably may be the fast-developing field of *Behavioral Law & Economics* whose origins can be found in Sunstein et al.'s foundational article on the topic in 1999.[57] Yet, so far, the field has focused on consumer law, financial regulation and *nudging* while other areas and actors within the

[56] *See*, e.g., Evan G. Williams, *The Possibility of an Ongoing Moral Catastrophe*, 18 ETHICAL THEORY & MORAL PRAC 971 (2015).

[57] *See* Cass R. Sunstein, Christine Jolls & Richard H. Thaler, *A Behavioral Approach to Law and Economics*, 50 STANFORD LAW REVIEW 1471-1550 (1998).

judicial system such as the judiciary[58] remain largely neglected. Because many of the cognitive biases explored by behavioral economists apply equally to lay-men and experts,[59] the strong present focus of behavioral law & economics on lay-men might be questioned.

Most interestingly for our purpose is the very recent emergence of the field of *experimental jurisprudence* which so far has primarily been focusing on empirical analysis.[60] However, given its embedding within the larger field of *experimental philosophy* which is intensively debating its normative implications, it seems fair to

[58] A very noteworthy exception to this phenomenon is the research conducted by Guthrie et al. *See*, *inter alia*, Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777–830 (2000), Andrew J. Wistrich et al., *Heart Versus Head: Do Judges Follow the Law or Follow Their Feelings?*, 93 TEX. L. REV. 855 (2015); Andrew J. Wistrich et al., *Judging the Judiciary by the Numbers: Empirical Research on Judges*, 13 ANNUAL REVIEW OF LAW AND SOCIAL SCIENCE 203 (2017); Andrew J. Wistrich et al., *Implicit Bias in Judicial Decision Making, How It Affects Judgment and What Judges Can Do About It*, *in* ENHANCING JUSTICE: REDUCING BIAS 87–130 (Sarah E. Redfield ed., 2018). *See* also Christoph Winter, *The Value of Behavioral Economics for EU Judicial Decision-Making*, GERMAN LAW JOURNAL (forthcoming, 2019).

[59] *See*, for instance, William Meadow & Cass Sunstein, *Statistics, Not Experts*, 51 DUKE L. J. 629–646 (2001); Daniel Kahneman & Amos Tversky, *Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment*, 90 PSYCHOLOGICAL REVIEW 308 (1983); Michael J. Roszkowski & Glen E. Snelbecker, *Effects of "Framing" on Measures of Risk Tolerance: Financial Planners are Not Immune*, 19 JOURNAL OF BEHAVIORAL ECONOMICS 237–246 (1990); for further sources, *cf*. the comprehensive list including military leaders, engineers, accountants, doctors, real estate appraisers, option traders, psychologists and lawyers presented in Guthrie et al., *supra* note 58, at 783.

[60] A first workshop on the topic was organized by Roseanna Sommers at Yale University in 2017.

assume that more attention will be paid to normative questions and underlying legal theories in the future.[61] As there is no reason to assume that the brain is using a special process to solve issues with regards to criminalization theory, legal philosophers can draw on the vast research outputs already gathered by (moral) psychologists and neuroscientists. At the same time, it is almost certain that further experiments specifically designed to analyze human intuitions about legal theories will not cast doubt on the general understanding of the brain's functioning even though they may shed some light on more specific applications of human reasoning.

3) Dual-Process Theory of Learning and Decision-Making

To make the knowledge about human decision-making processes more accessible, a large number of analogies have been introduced. The one most widely used might be the comparison to visual illusions, specifically the *Müller-Lyer Illusion*. Already Henry Sidgwick whose work is widely acknowledged among philosophers but almost completely overlooked by moral psychologists stated in the early 20th century:

[61] One reason for the current lack of legal normative analysis might be that the field is somewhat dominated by researchers with backgrounds in psychology and philosophy while concrete policy suggestions often require a larger understanding of the legal (procedural) landscape. While I assume that it is only a matter of time until legal theorists will join and benefit the debate, I expect this to happen much faster within common law jurisdictions due to more interdisciplinary education systems and less dogmatic approaches. Accordingly, the harm principle and legal moralism might profit from future insights from *experimental jurisprudence* significantly more than the theory of legal goods which is more common in civil law jurisdictions even though the *Rechtsgutslehre* could, in principle, equally profit from the behavioral approach.

> „I wish therefore to say expressly, that by calling any affirmation as to the rightness or wrongness of actions „intuitive“, I do not mean to prejudge the question as to its ultimate validity, when philosophically considered: I only mean that its truth is apparently known immediately, and not as the result of reasoning. I admit the possibility that any such ‘intuition’ may turn out to have an element of error, which subsequent reflection and comparison may enable us to correct; just as many apparent perceptions through the organ of vision are found to be partially illusory and misleading: indeed […] I hold this to be to an important extent the case with moral intuitions commonly so called.”[62]

Sidgwick, of course, did not have the modern means such as functional magnetic resonance imaging (fMRI) to test his theory. However, a century later, the comparison to visual illusions is still frequently used not only to indicate the limitations of human intuition but also to emphasize the ability to reflect and override these intuitions, if they turn out to be mistaken. Unfortunately, it is inherently more difficult to check whether one holds correct moral intuitions than whether one can trust her visual perception as the latter can easily be measured whereas the former must first overcome metaethical and epistemological debates in order to know how to know whether one holds a defensible moral position.

[62] HENRY SIDGWICK, METHODS OF ETHICS 211 (7th ed., 1907).

To simplify the distinction between intuition and deliberation, the thought processes are routinely referred to as system 1 (S1) and system 2 (S2) thinking. The difference between S1 and S2 thinking lays the foundation for the so-called *dual-process theory*. Joshua Greene, who applies dual-process theory to moral judgments, argues that the brain's operations may be best understood by an analogy to a digital camera.[63] While intuitions (S1) work like a camera's automatic settings such as "landscape mode", "night mode" and "portrait mode", deliberative thinking (S2) operates like "manual mode", in which all relevant settings can be adjusted by hand.[64] Indeed, intuitions, just like the automatic settings of a camera, can be very useful concerning routine every-day tasks given their efficiency.[65] They often lead to reasonable results and do not require a lot of navigation and adjustments.[66] On the other hand, manual mode is much more flexible and able to deliver remarkable results even in circumstances in which automatic settings would produce insufficient or misleading outcomes. Consequently, neither S1 nor S2 thinking is inherently better than the other, rather each of them has complementary strengths and weaknesses depending on the present situation.[67]

---

[63] GREENE, *supra* note 7, at 132-146; Joshua Greene, *Beyond Point-and-shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics*, 124 ETHICS 696-697 (2014).

[64] Joshua Greene, *The Rat-a-gorical Imperative: Moral Intuition and the Limits of Affective Learning*, 167 COGNITION 70 (2017).

[65] *Ibid*.

[66] *Ibid.*

[67] *Ibid.*

In order to further understand human decision-making, it is necessary to outline two shortcomings of the analogy. First, Greene acknowledges that human intuitions are constantly being updated through learning whereas the camera's automatic settings are fixed.[68] A better analogy might therefore be a "smart" camera, which updates its automatic settings based on some feedback mechanism such as what photos are kept or deleted.[69] A second problem is that the analogy does not capture the risk of *rationalizations*. Even if humans rely on S2 thinking, one still does not know for which purpose S2 has been used as it is neutral regarding the preferred consequences. For example, one may intuitively think that action ϕ ought to be criminalized and then use S2 thinking in order to rationalize the (misleading) intuition about ϕ rather than using S2 to analyse whether ϕ ought to be criminalized or not. Accordingly, deliberation through S2 may merely be used to rationalize intuition, a process Greene refers to as *intuition chasing*.[70] This is why one should be careful about basing the legitimacy of criminal laws on very abstract principles or requirements such as "interest" or "harm" as these terminologies allow rationalizing unreliable moral intuitions about specific actions by reference to broadly constructed interests such as the concern for public safety. Despite these minor shortcomings, the camera analogy makes clear that intuitions are not always reliable. They may be very efficient, and it may even be a rational choice to consult them

[68] *Ibid.*

[69] *Ibid*.

[70] Greene, *supra* note 63, at 718; *Cf. also* Joshua Greene, *The Secret Joke of Kant's Soul*, *in* MORAL PSYCHOLOGY, VOL. 3: THE NEUROSCIENCE OF MORALITY 35-79 (Walter Sinnott-Armstrong ed., 2007); *see* also Jonathan Haidt, *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, 108 PSYCHOLOGICAL REVIEW 814-834 (2001).

when it comes to everyday decisions, but one should be aware of the fact that automatic settings may not always deliver the best results. This observation is especially important when, as with criminalization, there is no time-limiting need to apply S1 thinking alone and the consequences of a wrong decision are severe.

4) The Arguments from Cognitive Biases and the Theory of Evolution

So far, I have argued that moral intuitions are not always reliable by referring to historical examples and the dominating view within the behavioral sciences. Yet, in order to draw any valuable information for theories of criminalization from this research, we need to understand under what conditions intuitions are more or less reliable and what the normative implications of such findings are. At least three potential arguments come to mind, which I will address in the following. I will briefly sketch out the implications of cognitive biases and evolutionary theory before I will examine the implications of dual-process theory in more detail in section III.

First, one may draw on the vast cognitive biases literature produced by experimental psychology and behavioral economics to argue that a specific bias may apply when thinking about the criminalization of a specific conduct and that related policies should take this into account. Let us call this the *argument from cognitive biases.* For instance, even if one favours the abstract principle of valuing future life just as much as present life, it needs to be considered that humans are particularly bad at thinking impartially

across time[71] and large numbers,[72] which in turn, is likely to influence the (intuitive) evaluation of criminal environmental protection laws given its (partial) justification of protecting a *large number* of *future* lives.

Secondly, instead of focusing on specific heuristics and biases one might zoom out and evaluate intuitions from an evolutionary psychological perspective. The so-called *evolutionary debunking arguments* (EDA) have led to an intensive discussion in moral philosophy over the past decade following Peter Singer's paper on "Ethics and Intuitions" in 2005.[73] The basic argument is as follows: An intuition ϕ is identified as likely to be influenced by evolutionary forces. It is stated that our best understanding of evolution indicates that it does not select for moral truth. Consequently, one ought not to trust ϕ with regards to its moral validity. A classic example for evolutionary debunking arguments from the moral psychological and philosophical literature is the case of Mark

[71] *See*, e.g., Ted O'Donoghue & Matthew Rabin, *Doing It Now or Later*, 89 AMERICAN ECONOMIC REVIEW 103-124 (1999).

[72] Joshua Greene & Jonathan Baron, *Intuitions about declining marginal utility*, 14 JOURNAL OF BEHAVIORAL DECISION MAKING 243-255 (2001).

[73] Peter Singer, *Ethics and Intuitions*, 9 THE JOURNAL OF ETHICS 331-352 (2005). Singer's work was again inspired by Joshua Greene's "THE TERRIBLE, HORRIBLE, NO GOOD, VERY BAD TRUTH ABOUT MORALITY, AND WHAT TO DO ABOUT IT" (Ph.D. dissertation, Department of Philosophy, Princeton University, 2002) and further developed in Katarzyna de Lazari-Radek & Peter Singer, *The Objectivity of Ethics and the Unity of Practical Reason*, 123 ETHICS 9-31 (2012); *see* also the related debate on the objectivity of ethics beginning with the widely discussed "Darwinian Dilemma" put forward by Street in Sharon Street*, A Darwinian Dilemma for Realist Theories of Value,"* 127 PHILOSOPHICAL STUDIES 109-166 (2006).

and Julie, adult siblings who decide to have sex with each other. Even though the case stipulates that they use protection and that the one-off encounter does not affect their healthy sibling relationship, many people believe that Julie and Mark act immorally.[74] Given that an intuitive aversion to incest is advantageous from the evolutionary point of view, it is reasonable to assume that evolutionary forces have either directly or indirectly selected for this intuition.[75] Needless to say, this is a controversially discussed issue within evolutionary biology and the development studies and uncertainty remains as to the roots of the incest aversion. The point I am making, however, is not that evolution is *the* one and only factor determining the aversion to incest, but solely that this is *a* plausible factor. From the perspective of evolutionary debunking arguments, the higher the likelihood of a significant influence of evolutionary forces on human intuitions is, the less reliable these intuitions are.

Despite the fact that the Mark-and-Julie case and Jonathan Haidt's related empirical research have been intensively debated in moral philosophy, they have been neglected by legal theorists. The German Ethics Commission on the prohibition of incest concerning § 173 German Criminal Code, for instance, even introduces Haidt's research but then fails to discuss its potential normative implications in an otherwise thorough analysis.[76]

---

[74] *See* Haidt, *supra* note 70.

[75] Debra A. Lieberman, John Tooby & Leda Cosmides, *The Architecture of Human Kin Selection*, 445 NATURE 727-731 (2007).

[76] DEUTSCHER ETHIKRAT, STELLUNGNAHME ZUM INZESTVERBOT 43-45 (2014), https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-inzestverbot.pdf.

This is not to argue that the EDA in this specific case or more generally upholds, but rather to outline that there are a lot of low-hanging fruits for legal philosophers to pick. This is especially so because the application towards criminalization theories is not as straight-forward as it might seem at first sight through the moral philosophical lens. In fact, even opponents of EDAs who argue that it is *unlikely* but not impossible that one can (indirectly) draw any normative implications from the sources of human intuitions might come to the conclusion that this reasoning casts enough doubt to argue in favor of decriminalization. In this regard, one might argue that EDAs are not sufficient to declare the moral intuition in question as *fully* unreliable, but simply as *less* reliable. As less reliable intuitions still have some value, proponents of this approach may conclude that incest should still be considered morally wrong, but acknowledge the *chance* that it is not wrong. This acknowledgment of uncertainty can lead the same philosopher to argue in favor of decriminalization due to *EDA-induced uncertainty* while still holding that incest is likely to be morally wrong.

The argument from *EDA-induced uncertainty* in favor of decriminalization may become stronger, the more one can agree on the current status of overcriminalization. In case one accepts that EDAs can cast some doubt on the immorality of a certain conduct while stipulating that too much behavior is already being criminalized, then one surely has good reason not to criminalize conduct which may not even be immoral. Hence, EDA-induced uncertainty offers an alternative way or even principle from which theories of criminalization seeking to address the problem of overcriminalization might significantly benefit.

5) Preliminary Summary

We have now seen that theories of criminalization are vastly based on human intuitions, and that evidence from history and the behavioral sciences including dual-process theory indicates that these intuitions are not always reliable. I have then briefly sketched out two arguments which make use of the knowledge about human intuitions for theories of criminalization, namely the *argument from cognitive biases* and the *argument from the theory of evolution*. I will now turn to the third option in more detail: the normative implications of dual-process theory for criminalization.

III. TOWARDS A NORMATIVE DUAL-PROCESS THEORY OF CRIMINALIZATION

Once convinced that dual-process theory may offer some insights about theories of criminalization, the first question is *where* to start. Unfortunately, the answer is less obvious than the question. As Winter argued,[77] one might begin with the examination of very abstract intuitions, such as those concerning the purpose of the state or the understanding of "human interests" and "harms". In fact, starting from abstract principles and moving to more concrete positions over time would mirror the typical way philosophers choose to develop their ideas. The problem with this approach, however, is that moral, political and legal theories often remain abstract and offer little practical guidance. As I have described in section I of this article, this allows one to rely on (and

[77] WINTER, *supra* note 1.

rationalize) moral intuitions, which we later classified as S1 thinking. Instead of picking specific intuitions and analyzing them from a dual-process theoretic point of view, it may be beneficial to look at intuitions themselves more specifically and examine when they are reliable. In order to do so, we first need to know more about the operations of the dual-process brain (1). Thereafter, I will outline under what conditions S1 is unreliable (2) and why S2 may be counter-intuitive, but still the best available option (3). I will then apply the insights to the old debate on *mala in se* and *mala in prohibita* and introduce the *normative dual-process theory of criminalization* (4).

1) Model-free and Model-based Learning and Decision-Making

As we have already seen, dual-process theory in its very basic form distinguishes between intuitive and more deliberative thinking, the automatic and manual settings in Greene's camera analogy. Many different attributes have been ascribed to the two thought processes with S1 commonly described as fast, automatic, efficient, instinctive, emotional, frequent, stereotypic and unconscious and S2 as slow, effortful, rational, analytical, infrequent, logical, calculating and conscious.[78] Even though it is tempting to argue that S2 thinking is intrinsically *better* from a normative point of view given the large number of heuristics and biases attributed to S1, one should be careful with rejecting S1 thinking altogether and instead try to examine when its automatic settings are likely to be insufficient.

[78] KAHNEMAN, *supra* note 6, at 19-31.

Drawing on the computational neuroscientific literature, recent findings by Fiery Cushman[79] and Molly Crocket[80] suggest that S1 and S2 thinking may be best understood as *model-free* and *model-based algorithms* respectively, rather than focusing on the widely held distinction between "emotional" and "rational" processes. Model-based algorithms (S2) carefully analyze the environmental surroundings, their causal relations, and only then evaluate which action has the best outcome.[81] They investigate different options, their possible consequences and select the path which maximizes rewards. Hence, model-based thinking is driven by expected value. In contrast, model-free learning algorithms (S1) are informed by experiences and attach positive or negative values to actions based on the extent to which those actions have previously been rewarded.[82] Consequently, model-free thinking does not need to carry a causal model of the world as it does not make decisions based on outcomes but the value of actions. Notably, model-free thinking does not appear to value all past experiences equally. More specifically, its tendency to emphasize the most recent observations enable it to adapt to a changing world.[83]

---

[79] Fiery Cushman, *Action, Outcome, and Value: A Dual-System Framework for Morality*, 17 PERSONALITY AND SOCIAL PSYCHOLOGY REVIEW 273–292 (2013).

[80] Molly Crockett, *Models of morality*, 17 TRENDS IN COGNITIVE SCIENCES 363–366 (2013).

[81] Cushman, *supra* note 79, at 277.

[82] *Ibid*.

[83] Aaron Courville, Nathaniel Daw & David Touretzky, *Bayesian Theories of Conditioning in a Changing World*, 10 TRENDS IN COGNITIVE SCIENCES 294-300 (2006).

The operations of model-free thinking can be illustrated by a phenomenon coined *moral dumbfounding*[84] and its most frequently cited case of Mark and Julie's incest adventure which we previously discussed regarding the applicability of evolutionary debunking arguments. The key finding of Haidt's related study is that many people consider incest to be wrong but few can precisely say why. While participants at first justify their judgment with references to harmful outcomes, the potential for birth defects, regret, family shame, and so forth,[85] a large number of participants ultimately ends up with the statement "it's just wrong", when investigators tell them that the aforementioned factors have explicitly been ruled out in the case of Mark and Julie. Moral dumbfounding which has also been found in other cases such as eating a dead family pet, burning the national flag and in some versions of the trolley problem[86] may therefore be classified as a model-free phenomenon. The intuition that incest is wrong does not depend on a causal model and potential consequences, instead the action is considered intrinsically wrong regardless of its consequences.[87]

---

[84] Fredrik Björklund, Jonathan Haidt & Scott Murphy, *Moral Dumbfounding: When Intuition Finds No Reason*, 2 LUND PSYCHOLOGICAL REPORTS 6 (2000) define moral dumbfounding as "the stubborn and puzzled maintenance of a moral judgment without supporting reasons"; Jonathan Haidt & Matthew Hersh, *Sexual Morality: The Cultures and Emotions of Conservatives and Liberals*, 31 JOURNAL OF APPLIED SOCIAL PSYCHOLOGY 191-221 (2001).

[85] Björklund et al., *supra* note 84.

[86] *See* Fiery Cushman, Liane Young & Marc Hauser, *The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm*, 17 PSYCHOLOGICAL SCIENCE 1082-1089 (2006).

[87] Note that there is some doubt as to whether the incest taboo is a product of model-free learning, *see* Cushman, *supra* note 79, at 286.

This observation itself (which we will come back to shortly) does not stipulate any argument for or against the criminalization of incest and other likely products of model-free learning and decision-making. In fact, model-free algorithms have many advantages as illustrated by Greene's camera analogy. More specifically, relying on past experiences can be a helpful to evaluate the likely outcome of similar present and future scenarios. However, model-free algorithms are unreliable in two situations Greene more recently referred to as the "bad training" and "bad data problem".[88]

2) Training and Data Problems of Model-free Learning and Decision-Making

The *bad training problem* occurs when the action's evaluation system changes. If intuitions are trained by ascribing positive or negative values based on *Y*, then those intuitions become unreliable, if ones' evaluation criteria change from *Y* to *Z*. For instance, if humans, as suggested by Greene and many others, are at least partially trained to reap the fruits of cooperation within a "tribe", then actions may be assigned positive values which put the human's own tribe ahead of other tribes. In this way, the moral intuitions that solve the puzzle of cooperation within groups, create conflict between groups.[89] Such tribalism may not only select for racism, nationalism and xenophobia but also indirectly influence political views and even one's credence on purely factual topics such as the

[88] Greene, *supra* note 64, at 72-73.

[89] Because Greene argues that moral intuitions, i.e. model-free algorithms may overcome Hardin's infamous t*ragedy of the commons* (1954), he refers to the resulting inter-tribal conflicts as the *tragedy of common sense morality*. *See* GREENE, *supra* note 7, at 19-27.

scientific evidence on climate change.[90] However, if humans favor a global community rather than tribal tendencies which may not be suited to solve some of the problems of the 21st century including climate change, threats from artificial intelligence and bio-risk, then model-free algorithms which have been trained on an outdated goal-set are very unlikely to deliver promising results. They are unreliable from the point of view of any set of goals which diverges from the one used to train these intuitions over thousands of years.

Aside from an evaluative framework, model-free algorithms need sufficiently good data to form robust and reliable intuitions. Luckily, humans as a species have been around for quite some time and were able to collect massive data sets which can be passed on through culture and evolution while individuals add their own data via personal experiences to it. For instance, the evolution of an aversion against direct personal violence *may* be the result of many of such experiences and their negative evaluations. More particularly, the aversion comes with the benefit of avoiding conflict which often has bad consequences for the perpetrator and victim alike.[91] At the same time, the crucial point about model-free learning mechanisms is that the value is ascribed to the action independent of its consequences. Thus, the intuitive aversion against personal violence applies irrespective of the consequences of the specific case. Against this backdrop, Greene convincingly argues that model-free thinking might lead us astray in circumstances such as the famous

[90] Greene, *supra* note 64, at 74; more specifically, *see* Dan M. Kahan, *The Polarizing Impact of Science Literacy and Numeracy on Perceived Climate Change Risks*, 2 NATURE CLIMATE CHANGE 732–735 (2012).

[91] *Ibid*., at 75 with further references.

footbridge case in which the "usual relationships between actions and consequences are reversed."[92] Indeed, history suggests that model-free learning algorithms have not been informed by data about the consequences of pushing a person from a footbridge in front of a trolley to save a larger number of lives.

More generally, one can argue that whenever causalities between actions and consequences significantly change, intuitions produced by model-free learning and decision-making become unreliable as they are based on outdated data of causal relations. Case in point, if direct personal violence is ascribed a negative value for tens of thousands of years, then intuitions about artificially constructed hypothetical cases which are unlikely to have occurred in real-life, such as the *footbridge case*, or which actions typically led to very different consequences will draw on data which does not fit the current situation. In other words: we use the wrong tools to solve the right problems. A bike is great for commuting to work, but if we try to climb Mount Everest, we might be better off without it.

Once again, we can bring this back to the Mark-and-Julie case. If our intuitions have been fed with data about the unfavorable consequences of incest over tens of thousands of years, then it is likely that an aversion against such actions is developed. However, if the causal connections change and, among other things, genetic risks for off-spring are mitigated, then those intuitions become unreliable because they are informed by outdated data. This is not necessarily to argue that incest may not be morally bad or even good, but

[92] *Ibid*., at 76.

it is to say that we cannot rely on human intuition to determine its moral status or criminalization. Lastly, it should not be assumed that human intuition is unable to adopt. As mentioned earlier, there is good reason to assume that model-free learning algorithms take recent observations more heavily into account than those from 10.000 years ago. Yet, it is highly unlikely that suddenly changing causalities, for instance due to technological innovation, are *sufficiently* considered.

3) Model-based Thinking & the Problem of Counter-intuitive Consequentialism

From a dual-process point of view, the alternative to model-free is model-based learning and decision-making which we identified as consequentialist. Needless to say, consequentialism comes in different versions. One might aim at maximizing pleasure over pain as favored by hedonistic act-utilitarianism but one could also argue in favor of satisficing a variety of different values. Many consequentialist theories have been proposed and at least just as many counter-arguments have been formulated to defeat a consequentialist approach in principle or specific types of it, such as utilitarianism. Note, however, that the general claim that one thought process is better suited to solve a specific issue does not entail that its results are the moral truth philosophers have been searching for. It merely states that it is the best option available. If model-free thinking is likely to be unreliable, then model-based thinking is our best option to figure out what conduct should be criminalized - regardless of whether this supports our favorite legal philosophy or not.

Many of the arguments brought up against consequentialist theories and utilitarianism in particular draw heavily on intuitions about strange hypothetical thought experiments ranging from trolley cases and utility monsters,[93] to happy societies whose well-being either depends on a young girl's suffering[94] or slavery.[95] Yet, regardless of the strength with which we hold the related intuitions, they may simply trigger intuitions based on an outdated or irrelevant data set thereby leading philosophers and criminal law theorist astray. As Greene puts it:

> "For any action that feels terribly, horribly wrong because of its typical real-world bad consequences, one can always construct an unrealistic hypothetical world in which its consequences are artificially stipulated to be good. And if we are willing to trust such intuitions, trained up on unrepresentative data, then our moral theorizing will inevitably be distorted."[96]

If model-based thinking (and consequentialism) would always be "intuitive", it would never conflict with model-free thinking and one could simply rely on currently held intuitions. However, this is improbable given that causal connections between action and consequences as well as training goals change over time. Furthermore, it seems unlikely

[93] ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA 41 (1974).

[94] Ursula K. Le Guin, *The Ones Who Walk Away from Omelas*, *in* NEW DIMENSIONS III 1-8 (Robert Silverberg eds, 1973).

[95] RAWLS, *supra* note 13, at 24-26.

[96] Greene, *supra* note 63, at 76.

that model-based thinking would have survived the pressure of natural selection in case it led to exactly the same outcomes as the significantly faster model-free thinking, thus, rendering it useless.

4) Mala in se & Mala Prohibita: *From Descriptive Dual-process Theory of Learning and Decision-making to the Normative Dual-process Theory of Criminalization*

So far, the analysis of model-free thinking has been focusing on its unreliability arising from data and training problems. However, causalities and training goals may also remain the same pertaining to specific actions over long periods of time. Direct personal force still leads to severe injuries and death just like it did in pre-historical societies and intuitions about the related need for criminalization (and punishment) may very well be reliable all things considered. In fact, one can observe that the consequences of specific conduct criminalized under (almost) any jurisdiction in time and space including murder and assault remained vastly the same.[97] Since there is no reason to assume that the causalities of such behavior have significantly changed, associated intuitions can be considered reliable enabling criminal law theorist to consult their model-free learning

[97] *Note*, however, that criminal laws on murder, manslaughter and assault throughout the world also apply in situations which do not involve direct personal violence and where causalities have changed significantly, e.g. with regard to drone strikes. When I refer to the abovementioned crimes as part of the traditional criminal law in the following analysis, I am only referring to conduct involving direct personal violence. The criminalization of drone strikes, for instance, should be understood through the lens of modern criminal law as introduced below even if the same (inter)national statute or case law applies.

algorithms successfully. Interestingly enough, if one takes a closer look at the type of conduct which consequences did not change over millennia, the question emerges whether dual-process theory can even shed new light on the old debate on *mala in se* and *mala prohibita*. I believe that it can.

The Latin expression *malum in se* is used to refer to conduct which is intrinsically wrong regardless of its legal regulation whereas *mala prohibita* are only wrong because of their prohibition. Often, legal philosophers use crimes which are considered to fall into the *mala in se* category like murder and assault to draw normatively relevant information for criminalization theories more generally.[98] I have to admit that, at first sight, this is a very appealing methodology. Since we seem to know that murder and assault are wrong, why not look at them more closely, examine what makes them wrong, and then try to figure out what other conducts are wrong? To illustrate this point, one might argue that because murder and assault are "wrong", all conduct which is to be criminalized ought to be "wrong" or at least entail some element of "wrongfulness".[99]

I argue that the consideration of dual-process theory may lead us to a different result. Indeed, I suggest that the distinction between *mala in se* and *mala in prohibita* may be best understood via the distinction between *reliable* und *unreliable* model-free S1 thinking. From this perspective, it is no coincidence that murder and assault are said to be intrinsically wrong and we can barely find anybody who would argue against their

[98] Dan Priel, *Criminalization, Legitimacy, and Welfare*, 12 CRIMINAL LAW AND PHILOSOPHY 657–676 (2018) refers to this as the "standard view" (658).

[99] *See supra* section I. 2) b).

criminalization. Related intuitions are likely to be the product of reliable model-free thinking which ascribes negative value to the action without further regard of its consequences. No wonder, criminal law theorists consider such crimes to be wrong independent of its consequences or legal regulation. Conversely, I am not aware of any *mala in prohibita*, which can be traced back to reliable model-free thought processes. On the contrary, many of the areas of so-called "modern criminal law" including, *inter alia*, commercial criminal law, environmental criminal law, medical criminal law, animal criminal law and competition criminal law regulate conduct which causal relationships dramatically changed over time. For instance, raising a pen to enter false information into a tax document which eventually shifts money from the state to a (wealthy) individual may have, all things considered, worse effects than raising one's fist to hit another person. However, the aversion against direct personal violence is significantly stronger than one's negative intuitions about tax evasion. The phenomenon that some individuals are so wealthy that their failure to pay taxes can (indirectly) have substantial effects on the well-being of thousands of other people is, meta-historically speaking, an extraordinarily recent development following the massive increase in GDP after the industrial revolution. At the same time, hitting another person has led to bad outcomes for as long as we know. Hence, the intuitive signals produced by our model-free algorithms might not be as strong in the tax evasion case as in the personal violence case even though its consequences may be worse.

Because model-free learning cannot draw from a sufficient data set regarding tax evasion and other modern criminal laws, the only other (and therefore best) option available to

evaluate the legitimacy of modern criminal laws can be found in model-based thinking.[100] Thus, when dealing with age-old criminal laws which are often referred to as *mala in se* such as murder and assault, one might reasonably refer to model-free or S1 thinking respectively. However, when examining the legitimacy of modern criminal laws on incest, environmental harm, tax evasion and the like, a consequentialist approach based on model-based or system 2 thinking respectively may be preferable - even and especially if its results are not as intuitively appealing as those produced by S1 thinking. Hence, the descriptive dual-process theory of learning and decision-making delivers valuable empirical insights which indirectly favor a *normative dual-process theory of criminalization* distinguishing between *mala in se* and *mala prohibita*, if we can agree that we should not consult unreliable thought processes. To summarize, the normative theory of criminalization states:

a) If we need to evaluate *mala in se*, that is to say traditional criminal law understood as dealing with familiar causalities between action and consequence, model-free thinking is likely to be reliable. Criminal law theorists may therefore consult their intuitions to evaluate the legitimacy of traditional criminal law.

b) If we need to evaluate *mala prohibita*, that is to say modern criminal law understood as dealing with new or changing causalities between action and consequence, model-free thinking is unlikely to be reliable. Instead, model-based

[100] Unless one would like to bet on a "cognitive miracle", Greene, *supra* note 64, at 745.

thinking is preferable and, hence, a consequentialist approach ought to be taken to evaluate the legitimacy of modern criminal law.

If the psychological and neuroscientific evidence on dual-process theory can be upheld,[101] and the normative implications for theories of criminalization can be defended, then the normative dual-process theory of criminalization has far-reaching consequences for criminal law theory. First of all, consequentialist approaches are much less represented in legal philosophy than in moral or political philosophy more broadly. Its implications may therefore be substantial. Secondly, modern criminal law as understood in this article covers large parts of today's existing criminal law in many jurisdictions as indicated by the previous non-exhaustive list of criminal offenses. Typically, intuitions about actions which are influenced by modern technologies or have inter-tribal, potentially even global, effects are particularly likely to be unreliable. Related present and future criminal laws should therefore be evaluated from a consequentialist point of view.

Although it is beyond the scope of this article to examine the variety of implications of the dual-process theory of criminalization, it should be pointed out that such an approach may not only have effects on the legitimacy of specific criminal offenses, but also on general principles such as the distinction between action and omission. Accordingly, humans consider it morally worse to harm a person actively than to passively allow one

[101] A recent and thorough study further indicates this, see Indrajeet Patil et al., *Reasoning supports utilitarian resolutions to moral dilemmas across diverse measures*, DOI 10.17605/OSF.IO/JDZFS (forthcoming, 2019).

to die, for example, by withholding a life-saving antidote.[102] As Cushman notes, such an effect can be explained by the fact that model-free value representations preferentially encode the value associated with the available actions but not with the ever-available option to omit action.[103] If an action leads to bad outcomes, the model-free learning processes preferentially encode "negative value for action" rather than "positive value for inaction."[104] In contrast, if an action is consistently rewarded, model-free learning preferentially encodes "positive value for action," rather than "negative value for inaction."[105] Hence, an omission cannot carry a forceful negative evaluation based on model-free learning and decision-making.[106] This may explain why many jurisdictions strongly distinguish between the criminalization of actions and omissions. However, if one accepts the view that omitting to help another person may be criminalized under certain conditions or if one aims at evaluating current criminal laws on omission, then model-free thinking might again be an unreliable tool given that humans intuitively do not assign any negative (or positive) value to omissions.

---

[102] Jonathan Baron & Ilana Ritov, *Omission Bias, Individual Differences, and Normality*, 94 ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES 74-85 (2004); Jonathan Baron & Ilana Ritov, *Protected Values and Omission Bias as Deontological Judgments*, *in* THE PSYCHOLOGY OF LEARNING AND MOTIVATION: VOL. 50. MORAL JUDGMENT AND DECISION-MAKING 133-157 (Daniel Bartels et al. eds., 2009); Cushman et al., *supra* note 86.

[103] Cushman, *supra* note 79, at 284.

[104] *Ibid*.

[105] *Ibid*.

[106] *Ibid*.

IV. CONCLUDING REMARKS

I have started with the analysis of current theories of criminalization focusing primarily on the widely endorsed *harm principle* and the *theory of legal goods*. Extensively interpretable notions such as "harm", "interest" or "legal good" allow intuitions rather than deliberative thinking to determine the legitimacy of criminal law. Consequently, I have argued in section II that a *behavioral approach to law and philosophy* more generally, and to criminalization in particular, is needed to understand which intuitions are reliable and under what circumstances our outdated hunter-gatherer thought processes lead us astray in a modern world. In short, I have outlined the need to have a look at the brain's manual, if we are serious about building a theory of criminalization which passes the test of time and will not be viewed by future generations as yet another tool for justifying moral atrocities.

Liberal approaches to criminalization have been somewhat at the center of my critique. As emphasized earlier, this was not because I generally disagree with the current mainstream liberal viewpoint in criminal law theory. On the contrary, for reasons I cannot possibly discuss in the remaining space, I consider the dual-process theory of criminalization to be in line with other liberal approaches. The reason for my critical remarks was rather that there is wide agreement that liberal theories offer the most limiting approaches to criminalization. If even they cannot effectively limit the impact of human intuition, what theories can?

I have then suggested three strategies on how the findings of the behavioral sciences may be incorporated into the legal philosopher's argumentative toolkit, namely the *argument from cognitive biases*, the *argument from the theory of evolution*, and ultimately analyzed the implications of the dual-process theory of learning and decision-making. Since dual-process theory and its particular interpretation adopted here distinguishing between model-free and model-based algorithms is not without its critics, the second section of this article might also be read as a future research agenda, which will hopefully encourage others to evaluate the (normative) implications of the behavioral sciences for legal philosophy and theories of criminalization irrespective of whether they consider dual-process theory as promising as I do.

The final section focused on introducing the *normative dual-process theory of criminalization*. While the dual-process theory of learning and decision-making as advocated for in the cognitive sciences is descriptive, dual-process theory of criminalization is normative. Accordingly, the analysis of the legitimacy of criminalization depends on whether traditional or modern criminal laws are examined. The latter which I take to be the same as *mala prohibita* should not draw on S1 thinking as model-free algorithms are unreliable in such circumstances due to outdated or insufficiently available data. Instead, model-based thinking which we classified as consequentialist is preferable for analyzing the legitimacy of modern criminal law. However, traditional criminal law provisions such as murder and assault may still be evaluated with the prevailing intuitive approach. In this vein, I also argued that it is no coincidence that crimes considered *mala in se* can be explained by reliable model-free

thinking.[107] Indeed, theories of criminalization should not necessarily concentrate on the analysis of such crimes for it is unlikely that our respective intuitions will change anytime soon and disagreement will emerge. Neither might it be methodologically beneficial to draw on such crimes as the transferal of intuitions about one action to another is likely to cause trouble.

Throughout the analysis I have looked at different examples from incest to tax evasion and concluded with some remarks on the infamous distinction between action and omission. The purpose of the examples was not to offer concrete guidance and decisive arguments settling the debate once at for all or at least from a dual-process theoretic perspective, but rather to outline the importance of taking the cognitive sciences into account when evaluating existing or potential criminal laws. My comments should rather be viewed as conversation starters than definitive answers. In fact, there are many crucial topics which I did not cover and which future research endeavors will have to deal with, such as what specific consequences one should aim for within model-based thinking.[108]

While it seems obvious that findings about human thought processes can offer valuable insights for all researchers, legal scholars have been particularly hesitant to engage with

[107] Conversely, the fact that the very actions which would be assigned a negative value from reliable model-free thought processes are considered *mala in se* by legal philosophers may even constitute a further argument in favor of understanding human decision-making processes as model-free and model-based algorithms.

[108] GREENE, *supra* note 7, at 190-210 favors happiness as a common currency, ultimately leading to classic utilitarianism he prefers to call *deep pragmatism* (289).

the behavioral sciences. Admittedly, there are important developments within *behavioral law & economics* and *experimental jurisprudence* more recently. However, the vast majority of this research has shied away from making behaviorally informed *normative* arguments. In fact, to most legal theorists such a proposal will still sound radical even though research by Greene, Haidt and others, from which this work enormously profited, has been gaining prominence within political philosophy for the past decade. More generally, law and legal philosophy are still lagging behind in the use of interdisciplinary approaches compared to other fields. This may be less in common law systems which have proven be more accessible to other fields than civil law jurisdictions. Nevertheless, findings about the functioning of the human brain may ultimately become an indispensable partner to both.