

Rationale-Based Defences in Criminal Law

– ‘I did it. I chose to do it. And here’s why...’ –

Mark Dsouza



• H A R T •
PUBLISHING

OXFORD AND PORTLAND, OREGON

2017

Hart Publishing

An imprint of Bloomsbury Publishing Plc

Hart Publishing Ltd
Kemp House
Chawley Park
Cumnor Hill
Oxford OX2 9PH
UK

Bloomsbury Publishing Plc
50 Bedford Square
London
WC1B 3DP
UK

www.hartpub.co.uk
www.bloomsbury.com

Published in North America (US and Canada) by
Hart Publishing
c/o International Specialized Book Services
920 NE 58th Avenue, Suite 300
Portland, OR 97213-3786
USA

www.isbs.com

**HART PUBLISHING, the Hart/Stag logo, BLOOMSBURY and the
Diana logo are trademarks of Bloomsbury Publishing Plc**

First published 2017

© Mark Dsouza 2017

Mark Dsouza has asserted his right under the Copyright, Designs and Patents Act 1988
to be identified as Author of this work.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means,
electronic or mechanical, including photocopying, recording, or any information storage or retrieval system,
without prior permission in writing from the publishers.

While every care has been taken to ensure the accuracy of this work, no responsibility for loss or damage occasioned
to any person acting or refraining from action as a result of any statement in it can be accepted by the authors,
editors or publishers.

All UK Government legislation and other public sector information used in the work is Crown
Copyright ©. All House of Lords and House of Commons information used in the work is Parliamentary
Copyright ©. This information is reused under the terms of the Open Government Licence v3.0
(<http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3>) except where otherwise stated.

All Eur-lex material used in the work is © European Union, <http://eur-lex.europa.eu/>, 1998–2017.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

ISBN: HB: 978-1-50990-295-8
ePDF: 978-1-50990-297-2
ePub: 978-1-50990-296-5

Library of Congress Cataloging-in-Publication Data

Names: Dsouza, Mark, author.

Title: Rationale-based defences in criminal law : "I did it. I chose to do it. And here's why..." / Mark Dsouza.

Description: Oxford ; Portland, Oregon : Hart Publishing, 2017. | Based on author's thesis
(doctoral - University of Cambridge, 2014) issued under title: A theory of rationale-based defences. |
Includes bibliographical references and index.

Identifiers: LCCN 2016056275 (print) | LCCN 2016057838 (ebook) | ISBN 9781509902958
(hardback : alk. paper) | ISBN 9781509902965 (Epub)

Subjects: LCSH: Justification (Law)

Classification: LCC K5086 .D76 2017 (print) | LCC K5086 (ebook) | DDC 345/.04—dc23

LC record available at <https://lccn.loc.gov/2016056275>

Typeset by Compuscript Ltd, Shannon

Printed and bound in Great Britain by TJ International Ltd, Padstow, Cornwall

To find out more about our authors and books visit www.hartpublishing.co.uk. Here you will find extracts,
author information, details of forthcoming events and the option to sign up for our newsletters.

INTRODUCTION

This is a study of the philosophical foundations of supervening defences in the criminal law. The term ‘defence’ is used by practitioners to refer to any claim that protects the defendant from punishment. However, different claims preclude punishment by operating at different stages of the criminal adjudication process. One category of defensive claims denies the defendant’s status as a responsible moral agent at the time of the offence. This category includes claims of infancy, automatism, insanity and intoxication. Another category works by disputing the existence of the actus reus or mens rea elements of the offence. Examples include claims of alibi or (depending on the mens rea stipulation in an offence) inadvertence. Yet another set of punishment-blocking claims invokes procedural objections against trying the defendant. This set includes claims relating to limitation, diplomatic immunity, *autrefois acquit* and *autrefois convict*. All of the foregoing claims, while defensive in a practical sense, are not defences in the sense relevant to this study. This study focuses on those substantive defences that operate once the prosecution has discharged its initial probative burden and has proved beyond reasonable doubt that the defendant is responsible *prima facie* for committing the offence charged. Defences that operate at this stage ‘block the presumptive transition from responsibility to liability’, without denying the defendant’s responsibility, by claiming that further relevant factors should block liability.¹ I refer to these as supervening defences, and include in this set of defences claims of duress, necessity, self-defence and the like. This set of supervening defences is traditionally subdivided into the categories of justification and supervening excuse. In this book, I will argue that all supervening defences exculpate based on the reasons for which the defendant chose to commit what was *prima facie* an offence, and therefore that all supervening defences are rationale-based defences.

The philosophy of justifications and excuses in the criminal law has been studied for centuries. In that time, several different and contradictory models of justification and excuse have been proposed and passionately defended. Of course, there is no preordained sense in which the terms ‘justification’, ‘excuse’ and their cognates must necessarily be used—these are not pre-existing logical categories that we discover, but rather categories that we invent in the process of using and describing them. So how does one go about proposing a model of justification

¹ RA Duff, *Answering for Crime* (Oxford, Hart Publishing, 2007) 263. See also AP Simester, ‘Mistakes in Defence’ (1992) 12 *OJLS* 295, 296.

and excuse, and how does one evaluate the strengths of a model or compare rival models? Most theories of justification and excuse appear, at heart, to be derived by inductive reasoning, and that is a useful way of approaching the problem. A theorist poses to herself a succession of questions about the outcomes that intuitively she would like her model of justification and excuse to generate. For instance, she may ask herself the following (and more) questions to plot her preferred outcomes:

1. whether the entitlement to supervening defences (or any of them) depends on demonstrating that, 'all things considered', no harm was done;
2. whether the availability of a supervening defence turns on the existence of an actual need for defensive action or on the defendant's perception that defensive action is required;
3. whether a person taking justified or excused action is required first to retreat if possible, even when the threat she faces is wrongful or undeserved;
4. whether a person taking defensive action should compensate the people (or any class of them) affected by such action;
5. whether a person is required to submit to justified or excused action; and
6. whether and, if so, to what extent the classification of a defence as a justification or an excuse affects the set of executive actions that the state is permitted to perform.

Having identified her intuitively preferred responses, she would then construct a theory of justification and excuse that can generate all or most of her preferred outcomes and yet be internally consistent. The outcomes that cannot be consistently accommodated within the preferred theoretical model of justification and excuse may then be rejected in favour of some other less preferred (but still plausible) outcome.

Such a theory of justification and excuse could be critiqued internally, on the basis that it is internally incoherent or contradictory, or externally, on the basis that the intuited answers that the theorist has chosen to accommodate in her theory are themselves implausible. An internal critique offers value by checking the internal logic, but a theory that successfully repels all such critiques still appeals only to those who share the theorist's moral intuitions. An external critique, on the other hand, argues across the model critiqued. If there is a genuine disagreement between the intuitions of the theorist and the critic, then neither can hope to convince the other.

But of course our moral intuitions are more nuanced than this sketch implies. Even if we feel that 'P' is the correct response to one of our outcome-plotting questions, we may still be willing to accept that 'Q' and 'R' are also plausible solutions, even if they are not the solution to which we are most drawn. Hence, it is possible for a person to believe that there are two or more models of justification that are plausible, and it may be possible for a theorist to agree with her external critic, and vice versa, that both the theorist's and the critic's models are plausible, and perhaps that they have different areas of relative strength and weakness. Even so, on balance we may prefer one theory to another, because on the whole we find the

intuitions accommodated by one more plausible than those accommodated by the other. And although different people may well form different judgements about which model they prefer, an internally coherent model that can accommodate a greater number of plausible outcomes *and* offer plausible explanations for either rejecting other intuitions or addressing the concerns underlying them in some other way would be especially persuasive. If, in addition, such a model is able to show how its rules link to fundamental and basic intuitions about the nature and purposes of the criminal law, then its claim to being the most useful way to conceptualise justifications would be even more convincing.

All of this is very abstract, but it has a bearing on the methodology and the arrangement of the argument that will appear in this book. Part I sets up the overall shape and scope of the inquiry by outlining the current approaches to theorising justification and supervening excuse, and considering an alternative approach to performing the same task. I start Chapter 1 by describing a hypothesis that almost all criminal law theorists explicitly or implicitly rely upon in order to set up their models of the criminal law in general, and the structural difference between a justification and an excuse in particular. I call this the ‘wrongness hypothesis’. Although the wrongness hypothesis is eminently plausible, I argue that it is fundamentally an unprovable assumption based on an intuited view of the structure of the criminal law, and not a conclusion in itself. Moreover, there are good reasons to doubt the usefulness and appropriateness of the wrongness hypothesis approach for the criminal law. Therefore, I rely on some of the other literature on this subject to briefly describe an alternative account of the manner in which the criminal law functions, and I draw out the implications of this model for the divide between justifications and supervening excuses. Specifically, I argue that under this model, the difference between justifications and supervening excuses is the *quality* of the defendant’s reasons for committing a *prima facie* offence. I call this the ‘quality of reasoning hypothesis’. This hypothesis too is based on an unprovable assumption—one that incidentally entails the conclusion that the wrongness hypothesis, at least in the form that I describe it, is incorrect. Yet, because the broad outcomes that it generates are at least plausible at first blush, the quality of reasoning hypothesis cannot be dismissed out of hand. Therefore, I set out to explore it in greater detail.

Part II of this book considers what a criminal legal system in which defences were arranged based on the quality of reasoning hypothesis would look like. I start in Chapter 2 by expanding on the proposed alternative model of how the criminal law works. In particular, I describe this model’s implications for disputed theoretical issues, such as the way offence definitions ought to be framed and read, and the way we should conceive of culpability and criminal blame. Admittedly, not all of these implications track the conclusions of (some versions of) the more mainstream theories of the criminal law. However, I show that there is sufficient academic support for each of them to suggest that they are at least plausible. Even though neither overall model of the criminal law has a knock-down argument to defeat the other, in the chapters that follow I set out to demonstrate that the

alternative model being set up herein offers a preferable account of criminal law. I do so on the basis that it generates outcomes that better track our moral intuitions, especially in respect of putatively justified agents and agents who find it difficult to meet criminal law standards due to learning difficulties and other mental conditions not amounting to insanity.

I begin that endeavour in Chapter 3 by identifying certain propositions that I take to be axiomatic and foundational for a modern system of criminal law. Substantively, I take the state's commitment to liberal principles of non-interference to be axiomatic, and therefore argue that the core of an *ideal* state's criminal law cannot contain rules that regulate matters purely in the private domain. Structurally, I take it to be axiomatic that at least the core of the criminal law in a rule of law state is related to morality. In other words, subject to certain caveats and limitations, the criminal law at core endorses and enforces moral norms.² I show that most moral theories build upon an understanding of 'the moral good' that is in some way contingent upon the continued survival of humans. Their axiomatic value statements only make sense if they presuppose the existence of beings that are recognisably human. This makes the ontological identity of humans prior to, and outside the content of, a human morality, and of any normative system, including that of the criminal law, which derives from it. I identify from existing accounts of human nature certain features that are relevant to the operation of criminal law defences and accepted as being constitutive of a generalised instance of the species of human beings. I treat these as setting the main logical limits or boundaries of the guidance that can be offered by a moral criminal law in respect of defensive or self-protective conduct. Next, I describe the set of liberal norms that must exist to support 'constituent rights'—rights created when the state clothes the aforementioned limits of its power with the trappings of a legal entitlement. In doing so, I describe a special immunity from blame that must be recognised in respect of actions arising out of constitutively human behaviour, and argue that the criminal law defence made available for actions taken to defend against threats to life, major physical injury and the like is best understood as a translation into law of this special moral immunity from blame. I then extend my analysis to consider 'posited rights'—entitlements entirely contoured by the state in the exercise of its political authority. Using the norm structure described in relation to constituent rights as a template, I describe one set of norms that might plausibly be constructed to support posited rights. I propose that, cumulatively, the normative structures for supporting constituent rights and posited rights constitute a sufficiently (for my purposes) detailed model of the system of norms that underlies the criminal law, even though it addresses only directly victimising acts.

Part III of this book focuses on demonstrating how the aforementioned propositions about the underlying structure of the criminal law translate into

² Perhaps a better way of describing this move is to say that I choose to restrict my audience to those who share my commitment to core liberal ideals and have a morally distinctive conception of the criminal law, and then set out to describe a model that will appeal to this audience.

recognisable doctrinal rules. In Chapter 4, I explain how the system of conduct norms outlined in Chapter 3 maps onto the doctrinal rules of a modern criminal law. In doing so, I suggest that the paradigmatic forms of justified conduct conform to guidance contained in a certain subset of conduct norms from the criminal law's underlying system of norms and, as such, paradigmatic justifications guide conduct. By contrast, paradigmatic rationale-based excuses, which we do not expect to guide conduct, do not mirror any part of the conduct guidance in the criminal law's underlying system of norms. I also argue that necessity should be understood as a non-paradigmatic form of justification, in that although we expect this justification to guide future conduct, we cannot trace it to any of the criminal law's underlying conduct norms.

Chapter 5 is a more detailed study of the forms of justification that I have identified as being paradigmatic. I set out the scope of these justifications and make arguments about how they should be applied in the criminal law. In particular, I explain why I identify these defences as conforming to the paradigm of a justification, make arguments as to the perspective that ought to be adopted when evaluating a person's claim to a paradigmatic justification, and explore the limits of these justifications.

I focus on rationale-based excuses in Chapter 6 and build upon John Gardner's suggestion that we excuse conduct that conforms to societal normative standards for behaviour even though it falls short of the criminal law's normative standards. I do so by explaining how the notion of hypocrisy in blaming generates the normative pull of rationale-based excuses. I argue that although all rationale-based excuses are united by their underlying appeal to the principled avoidance of hypocrisy in blaming, the model of rationale-based excuses generated on this view is very flexible and therefore rationale-based excuses can look very different in different jurisdictions.

Thereafter, in Chapter 7, I argue that the defence of lesser-evils necessity can and should be conceived of as a special deeming justification, rather than as the paradigmatic justification. I propose a theory as to the source of the necessity defence's exculpatory pull and compare the operation and limits of this defence to those of paradigmatic justifications. I also apply this conception of lesser-evils necessity to many prominent English cases in which this defence has been considered (or might have been considered) in order to demonstrate the sorts of liability outcomes that it generates.

Finally, in Chapter 8, I describe the outcomes generated by this model of justification and excuse in relation to the outcome-plotting questions with which we began (along with a few other such questions). In doing so, I flesh out in greater detail the alternative model of justification and supervening excuse outlined in Chapter 2, and offer suggestions about how the intuitions that cannot be accommodated within this model may nevertheless be addressed by the state using one or more of the other tools of social coordination at its disposal. I argue that each of the outcomes generated by this model of rationale-based defences is plausible, even when it is not the current preferred mainstream response. To my mind of

course, these outcomes are more than plausible—I think that they are the most plausible responses to the outcome-plotting questions—but even someone who does not share my intuited preferences might be willing to accept that they are not implausible. If so, then provided that such a reader: (a) accepts the axiomatic assumptions with which I began; (b) broadly accepts the arguments used to reach these conclusions from those initial axioms; and (c) accepts that most conflicting outcome intuitions can be adequately addressed using other tools of governance available to the state, she has a very good reason to adopt my model of the distinction between a justification and a rationale-based excuse.

Even the most convincing philosophical theory relies on the reader being willing to suspend her disbelief long enough to let the roots of the argument take hold. Whether this theory is convincing or not is for the reader to decide, but it is probably true that a reader who is willing to accept that my starting points are at least plausible, even if not wholly appealing to her, has a far better chance of finding something of value in the pages that follow.

Part I

Overview

1

The Proposed Borders of Justification and Excuse

Many a theorist has tried to map the ‘perplexing borders of justification and excuse’,¹ but no consensus has emerged as to the specific features that differentiate these species of supervening defences. However, notwithstanding the cacophony of dissent, one assumption that is ubiquitous (albeit often unspoken) in the writings of most criminal law theorists is that justifications negate the wrongness of what occurred due to the prima facie criminal conduct, whereas excuses negate the defendant’s blameworthiness for engaging in the conduct. In this chapter, I critique this assumption and tentatively set out the contours of an alternative approach to distinguishing between justifications and excuses, which I will defend in the rest of this study.

1.1 The ‘Wrongness Hypothesis’

Despite disagreeing on points of detail, a large majority of theorists who have studied the justification/excuse distinction agree on the validity of what I call the ‘wrongness hypothesis’. Some background is necessary to understand the content of this hypothesis.

People like Fletcher,² John Gardner,³ Simester,⁴ Sendor⁵ and Husak⁶ agree that in order to be justified, an agent must be motivated by good subjective reasons

¹ K Greenawalt, ‘The Perplexing Borders of Justification and Excuse’ (1984) 84 *Columbia Law Review* 1897.

² GP Fletcher, ‘The Right Deed for the Wrong Reason: A Reply to Mr. Robinson’ (1975) 23 *UCLA Law Review* 293, 320.

³ J Gardner, ‘Justifications and Reasons’ in AP Simester and ATH Smith (eds), *Harm and Culpability* (Oxford, Clarendon Press, 1996) 104–05.

⁴ AP Simester, ‘On Justifications and Excuses’ in L Zedner and JV Roberts (eds), *Principles and Values in Criminal Law and Criminal Justice* (Oxford, Oxford University Press, 2012) 99.

⁵ BB Sendor, ‘Mistakes of Fact: A Study in the Structure of Criminal Conduct’ (1990) 25 *Wake Forest Law Review* 707, 766.

⁶ DN Husak, ‘Justifications and the Criminal Liability of Accessories’ (1989) 80 *Journal of Criminal Law and Criminology* 491, 516–17.

when committing what is prima facie an offence. They also insist that no action can be justified if it results in an outcome that is, all things considered, wrongful. At best, such an action may be excused. In his theory, Fletcher uses ‘wrongful’ as meaning ‘contrary to Right’, ie, contrary to the objective legal order, or the objective greater good.⁷ Similarly, for John Gardner⁸ and Simester,⁹ an outcome is considered wrongful if there is no non-excluded guiding reason to cause it. In other words, if in the circumstances that truly exist an agent ought not to have brought about a particular outcome, then she cannot be justified in bringing it about, even if she did so in the belief—however reasonable—that circumstances calling for her actions existed. Yet, they also require for justification that the agent act for one of the available guiding reasons. Sendor’s theory of criminal conduct employs the concept of ‘dual wrongfulness’, whereby a violation of the victim’s right is one half of wrongfulness, and disrespect for the rights of the victim is the other. A justification negates the first half of wrongfulness, whereas an excuse negates the second half. Thus, for Sendor too, a justification negates the presence of an objectively wrong outcome—the violation of someone’s rights. A person who mistakenly thinks she is justified may at best negate the second half of Sendor’s conception of wrongfulness.¹⁰ Similarly, Husak conceptualises wrongs in generally ‘objective’ outcome-related terms and treats justifications as the negation of this wrong.¹¹

While Eser subscribes to the proposition that a justification negates a wrong, he is less categorical about his conception of a wrong. Nevertheless, he seems to adopt Fletcher’s objective outcome-related conception thereof.¹² Gur-Arye has argued that the use of the justification/excuse distinction in the criminal law is unhelpful,¹³ but where she does use it, she too adopts a view that is in line with Fletcher’s view.¹⁴ Accordingly, both Eser and Gur-Arye also broadly accept that justifications negate the objective, outcome-dependent wrong, but add that the agent must commit the prima facie offence for subjectively appropriate reasons in order to be justified.

⁷ GP Fletcher, ‘Rights and Excuses’ (1984) 3(2) *Criminal Justice Ethics* 17. See also GP Fletcher, ‘The Right and the Reasonable’ (1985) 98 *Harvard Law Review* 949, 966–67.

⁸ Gardner, ‘Justifications and Reasons’.

⁹ Simester, ‘On Justifications and Excuses’ 99.

¹⁰ Sendor, ‘Mistakes of Fact’.

¹¹ Husak, ‘Justifications and the Criminal Liability of Accessories’ 494–96, 504, 509.

¹² A Eser, ‘Justification and Excuse’ (1976) 24 *American Journal of Comparative Law* 621, 621–23, 635, 637.

¹³ M Gur-Arye, ‘Should a Criminal Code Distinguish between Justification and Excuse?’ (1992) 5 *Canadian Journal of Law and Jurisprudence* 215.

¹⁴ M Gur-Arye, ‘Legitimizing Official Brutality: Can the War against Terror Justify Torture?’ (2003) dx.doi.org/10.2139/ssrn.391580, 21.

Robinson¹⁵ and Westen¹⁶ demur insofar as the agent's subjective reasons for acting is concerned. They argue that irrespective of the agent's reasons for acting, if the outcome the agent brought about was not objectively wrongful, she is justified in causing it (although she might still be convicted of an attempt to commit the crime that she *prima facie* committed).

Nevertheless, all of the scholars mentioned above are united in that while they are comfortable with excuses being analysed on a subjective basis, they either reject subjectivist bases for justifications altogether or require them to coincide with the negation of the objective, outcome-related 'wrongfulness' of what occurred. At the heart of each of their theories of justification and excuse, then, is the proposition that justifications negate the objective outcome-related 'wrongness'¹⁷ of the *prima facie* offence committed, whereas excuses relate to factors personal to the actor that negate the actor's blameworthiness for committing the *prima facie* offence.¹⁸ This is what I call the 'wrongness hypothesis'.

Uniacke and Duff also subscribe to the wrongness hypothesis, but they do so in a slightly different manner. Uniacke's theory is that whether conduct is justified rather than excused depends on the perspective from which it is being evaluated. For Uniacke, moral justification must be judged on the basis of the facts as the agent reasonably perceived them (or, as she terms it, 'agent-perspectively'). However, a legal justification is an all-relevant-things-considered judgment that it was factually appropriate to have engaged in the particular conduct.¹⁹ Hence, Uniacke would call a putatively justified agent 'perspectively justified' and would recognise that such an agent has a complete moral justification and a legal basis for a complete excuse. Duff uses slightly different labels—he would say that a putatively justified agent acts in a manner that was 'warranted' and thus morally justified. However, he insists that legal justification requires both warrant and objective

¹⁵ PH Robinson, 'A Theory of Justification: Societal Harm as a Prerequisite for Criminal Liability' (1975) 23 *UCLA Law Review* 266, 272–73. See also PH Robinson, 'Rules of Conduct and Principles of Adjudication' (1990) 57 *University of Chicago Law Review* 729, 749–50; and PH Robinson, 'The Bomb Thief and the Theory of Justification Defenses' (1997) *Criminal Law Forum* 387, 394–99.

¹⁶ P Westen, 'An Attitudinal Theory of Excuse' (2006) 3 *Law and Philosophy* 289, 306.

¹⁷ Note that the usage of the term 'wrongness' for the present purposes does not coincide with the sense in which I use the term 'wrong' in advancing my own thesis. For my own preferred usage, see note 22 in Ch 3 below.

¹⁸ Admittedly, this is a gross simplification. However, I believe that the features that each of these scholars associates with justification can be traced to this underlying (and sometimes unspoken) assumption, since these features would not follow if the stated proposition were false. See also RA Duff, 'Rule-Violations and Wrongdoings' in S Shute and AP Simester (eds), *Criminal Law Theory: Doctrines of the General Part* (Oxford, Oxford University Press, 2002) 62; MN Berman and IP Farrell, 'Provocation Manslaughter as Partial Justification and Partial Excuse' (2011) 52 *William and Mary Law Review* 1027, 1032. I might add that I do not think that subscription to the wrongness hypothesis is a necessary feature of any consequentialist, deontological or hybrid account of the ontology of a wrong. That is a further choice. I have identified the theorists named because they do make that further choice and I think that they represent the predominant line of thinking on this issue.

¹⁹ S Uniacke, *Permissible Killing* (Cambridge, Cambridge University Press, 1994) 12.

rightness.²⁰ As such, for the purposes of legal justification, though not moral justification, both Uniacke and Duff subscribe to the wrongness hypothesis.

The wrongness hypothesis is not universally accepted, of course. Greenawalt²¹ and Baron,²² for instance, theorise the availability of a justificatory defence broadly on the basis of subjective perceptions of facts, and would accordingly allow that an actor whose actions do occasion an objectively wrong outcome can be justified. Nevertheless, it is no exaggeration to state that the dominant explanation of the theoretical distinction between justifications and excuses relies upon the wrongness hypothesis.

1.2 Questioning the Wrongness Hypothesis

Theories premised on the wrongness hypothesis insist that a (legally) justified act does not occasion a legal wrong and that a ‘putatively justified’ actor who in fact authors a legal wrong cannot be justified in law. In these theories, the label ‘justified’ communicates (to the public and to the defendant concerned) a composite judgment about the actor and the action—the judgment that the actor did the right thing *and* that objectively the right thing happened. When either one of these statements is not true—ie, either the actor did not subjectively, and in the transitive verb sense, *do* the right thing, or she did, but all things considered the right thing did not happen—the ‘justified’ label is withheld.

Greenawalt’s contrary position is premised on his insistence that a ‘person may distinguish his evaluation of the desirability of the act from his evaluation of the actor’s conduct.’²³ He argues that:

So long as one exercises the best possible judgment on the facts he can reasonably acquire, the existence of other facts knowable only in some practically unimportant sense is immaterial for purposes of moral evaluation. His act is justified even if, in retrospect, the estimation turns out to have been wrong. Neither moral relevance nor ordinary usage supports distinguishing between justification and excuse on the basis of a line between unknowable and knowable facts.²⁴

This line of argument stems from a counter-definition of justification as being a matter of moral evaluation of the actor and not the act. Thus, in order for the

²⁰ RA Duff, *Answering for Crime* (Oxford, Hart Publishing, 2007) 281–84.

²¹ K Greenawalt, ‘Distinguishing Justifications from Excuses’ (1986) 49(3) *Law and Contemporary Problems* 89, 91–92, 102. See also Greenawalt, ‘The Perplexing Borders’ 1903.

²² M Baron, ‘Justifications and Excuses’ (2005) 2 *Ohio State Journal of Criminal Law* 387, 393, 396–98. See also M Baron, ‘The Provocation Defense and the Nature of Justification’ (2009) 43 *University of Michigan Journal of Law Reform* 117, 124–30.

²³ Greenawalt, ‘Distinguishing Justifications from Excuses’ 94. See also his comments at 102 of the same essay, where he explicitly rejects the wrongness hypothesis.

²⁴ *ibid* 94–95, footnotes omitted.

actor to have been justified, it is not necessary that the act be justified—ie, that the ‘correct’ outcome resulted from her actions. This proposition also finds support in the common law, according to which, if a person acts in self-defence on the basis of a genuine belief in facts that, if true, would justify her acting in self-defence, then she is justified, and has a complete defence to a crime of personal violence, even if it transpires that her belief was unreasonable.²⁵ Baron argues along similar lines.²⁶ Both, then, reject the wrongness hypothesis.

Which approach is correct? Is justification an evaluation of the merits of both the actor and the act or is it just an evaluation of the actor? Both approaches have some appeal and, to be clear, neither approach has a knock-down argument against the other, since both are premised on intuited understandings of what it is to be justified. However, there are good reasons to doubt the usefulness of the wrongness hypothesis approach. Wrongness hypothesis-based theories use the label ‘justified’ only when the label can also communicate to the public the finding that the right thing happened—that in some sense the victim has no legitimate grounds to complain about the violation of her rights. However, it does so at the cost of obscuring information about whether the actor *did* the right thing. It uses the same label for actors who *did* the right thing and actors who did not, in cases in which, all things considered, the right thing did not happen. Information about whether the victim deserved the rights violation she suffered or whether the breach of the regulatory provision served a good purpose may well be of interest to the public. However, one doubts that the criminal court’s judgments are the appropriate medium to convey this information, especially when they can only do so at the cost of information about the defendant’s personal deservingness of blame. Arguably, a criminal conviction or acquittal (whether by way of justification or excuse) should be a ruling about the defendant, because at least in the popular imagination, it assigns a label primarily to the defendant. Any information that it conveys about whether the right thing (all things considered) happened should be purely incidental,²⁷ and it certainly should not obscure information about the defendant. Intuitively, the sense one gets is that the criminal court’s primary purpose is to judge the defendant, and not to judge whether the victim deserved to suffer a violation of her rights. This latter task seems best left to civil courts, which are responsible for determining whether a victim deserves compensation.

However, a lot has been written about justifications and excuses on the basis of the wrongness hypothesis, and so we should also consider the normative

²⁵ *Beckford v R* [1988] AC 130, 144.

²⁶ Baron, ‘The Provocation Defense’ 124–30.

²⁷ In other words, the primary message in a criminal conviction should be ‘We have considered the facts of this case and conclude that D is a criminal—she committed ϕ offence’, not ‘We have considered the facts of this case and conclude that ϕ offence has been committed [against V]’.

argument for the appropriateness of the wrongness hypothesis. Since the term 'justification' has no preordained meaning, there is no 'correct' approach to theorising it that is simply waiting to be discovered. Hence, the argument for the most appropriate way to understand justification has to be normative, and the strength of any such argument depends on its rational appeal and the plausibility of the system it generates. The wrongness hypothesis is appealing because it is elegant and simple, and it seems to fit neatly into the structure that we associate with criminal law. The theory taps into our intuition that in the same way as offence definitions can be split up into elements of *actus reus* and *mens rea*, the categories of defences can also be split up into those that negate concerns underlying the *actus reus* stipulation, dubbed 'justifications', and those that negate concerns underlying the *mens rea*, dubbed 'excuses'.²⁸ This approach is attractive, but it makes an unsubstantiated logical leap in assuming that by conceiving of offences and defences symmetrically, we can formulate a good description of how criminal laws ought to function. In fact, to the best of my knowledge, there is no (non-intuited) reason that has been offered to substantiate the assumption that the *actus reus*/justification and *mens rea*/excuse symmetry has anything to do with normatively establishing the manner in which the law ought to function. True, offence and defence definitions (and, within them, *actus reus*, *mens rea*, justification and excuse stipulations) perform different functions within the criminal law, but they are not *how* the criminal law functions. At the very least, they do not found the only account of how the criminal law functions.

Another plausible account, and one that offers more narrative exposition, says that the criminal law functions by *ex ante* stipulating conduct that should be avoided and *ex post* evaluating cases in which the conduct happens anyway in order to determine the blameworthiness of the actor.²⁹ On this account, offence stipulations (which set out the *actus reus* and *mens rea* requirements) relate to the *ex ante* stage of criminal law, and so their features would be adapted to the function that they perform, *viz* providing prior conduct guidance with a view to avoiding certain harms. On the other hand, supervening defences are raised after the harm has occurred, at the stage of trial, in order to protect the defendant against a blaming judgment. Arguably, then, their features should relate primarily to the function that they perform at this stage, *viz* undermining the apparent blameworthiness of the defendant. On this view, all supervening defences would function by negating (at least some aspects of) the defendant's blameworthiness, and none of them need necessarily negate the occurrence of a wrong.

One might object that on any plausible moral account of criminal law, there must be *some* moral value ascribed to what happened, such that when a

²⁸ See in this connection Eser's account of the evolution of this approach in German criminal law philosophy: Eser 'Justification and Excuse' 626–29.

²⁹ See, for instance, AP Simester and A von Hirsch, *Crimes, Harms, and Wrongs* (Oxford, Hart Publishing, 2011) 3.

morally wrong thing happens, some adverse moral judgement of the actor who caused it should follow, even if we decide to excuse the actor on account of her non-culpability. The alternative account I have just outlined seems unable to accommodate this moral value, since it disconnects the question of justification from whether a moral wrong occurred. However, this objection proceeds on the false assumption that the moral value of what happened must be accommodated at the *ex post* stage of determining the blameworthiness of an actor. There is no reason that this must necessarily be true and if it is not, then this creates the logical space necessary in the alternative account outlined above to ascribe moral value to what actually happened without necessarily connecting justification to the non-occurrence of a wrong. One alternative stage of the criminal process at which the moral value of what happened can be accommodated is the *ex ante* stage. One might, for instance, frame the *ex ante* stipulation meant to prevent a moral wrong 'φ' such that it is only violated when φ actually happens (as opposed to when the actor tries to cause φ or does not take care not to cause φ). In a system like this, when φ fortuitously did not happen despite the actor's best efforts, the stipulation against causing φ would not be violated and the actor would attract no blame *vis-a-vis* the stipulation against causing φ (although she may well attract blame *vis-a-vis* other stipulations). In taking account of the fact that what happened was not φ, this account of the criminal law would ascribe moral value to what actually happened.

Despite its widespread currency, then, the wrongness hypothesis has no argument to conclusively establish its superiority over an account of supervening defences that explains them by reference to the functions they perform at the *ex post* stage of criminal law. To be fair, the converse is also true. However, since there are good reasons to doubt the usefulness of the wrongness hypothesis, it may be time to consider whether an account of supervening defences that explains them by reference to the functions they perform at the *ex post* stage of criminal law is more convincing. To this end, I make the assumption, for the purposes of this study, that the wrongness hypothesis is incorrect or, at least, unhelpful. Instead, I describe rationale-based defences in terms of the stages of the criminal law. If the system of justification and excuse that I describe is more persuasive than those predicated on the wrongness hypothesis, then that would go some way towards vindicating my assumption.

1.3 Conduct Rules, Decision Rules and the Stages of Criminal Justice

Let us examine further the proposition that the criminal law functions in discrete temporal stages. In framing their theory about the role of deterrence in criminal justice, Simester and von Hirsch observe that the criminal justice system operates

in stages and that it has different functions at each stage.³⁰ The first stage of criminal justice is forward-looking or prospective. It has to do with the criminal law's purely norm-setting function. At this stage, the criminal law is concerned with ex ante proscribing harmful conduct. The second stage of criminal justice is adjudicatory and evaluative, or backward-looking. After a general criminal law norm is violated, the criminal justice system must evaluate the prima facie offender's conduct to see whether, in view of her conduct, she deserves a blaming judgment. If that is the case, it must also evaluate whether she deserves punishment and, if so, how much.

A closely related and overlapping—arguably even congruent—concept to which Dan-Cohen draws attention is the 'old but neglected idea that a distinction can be drawn in the law between rules addressed to the general public and rules addressed to officials'.³¹ The former set of rules are 'conduct rules' and the latter set are 'decision rules'. Dan-Cohen argues that any given rule may be a conduct rule, a decision rule or both.³² In general, he explains that conduct rules give notice to the public about conduct that is proscribed and subject to criminal sanction, whereas decision rules guide decision-makers like the police, prosecutors and judges in assessing the blameworthiness of, and imposing sanctions upon, conduct rule violators.³³ Conduct rules, then, are concerned with ex ante guiding behaviour, so as to prevent the occurrence of a particular harm, whereas decision rules are concerned with allocating blame appropriately after the harm has occurred.

Because the primary rationale of conduct rules is to give notice, it is argued that they should be intelligible to the public and must clearly stipulate legal requirements so that people can avoid committing criminal offences. Accordingly, they should be simple, brief and grounded in objective criteria in order to avoid unpredictability, vagueness and ambiguity.³⁴

Decision rules are addressed to officials empowered and trained to make judgments. They are routinely broad and open-ended, allowing the decision-maker to take into account the complex and varied situational factors of a particular case,

³⁰ Simester and von Hirsch, *Crimes, Harms, and Wrongs* 3. The authors divide the operation of substantive criminal law into three stages: criminalisation; adjudication and conviction; and punishment. This study is concerned with the first two stages. See also PH Robinson, *Structure and Function in Criminal Law* (Oxford, Clarendon Press, 1997) 125.

³¹ M Dan-Cohen, 'Decision Rules and Conduct Rules: On Acoustic Separation in Criminal Law' (1984) 97 *Harvard Law Review* 625. See also M Dan-Cohen, 'Reply' in PH Robinson, SP Garvey and KK Ferzan (eds), *Criminal Law Conversations* (Oxford, Oxford University Press, 2009).

³² Dan-Cohen, 'Decision Rules and Conduct Rules' 631.

³³ For a summary of Dan-Cohen's argument and a general overview of the conduct rule/decision rule idea, see MR Gardner, "Decision Rules" and Kids: Clarifying the Vagueness Problems with Status Offense Statutes and School Disciplinary Rules' (2010) 89 *Nebraska Law Review* 1, 4–7. See also Robinson, 'Rules of Conduct' 731; MN Berman, 'Justification and Excuse, Law and Morality' (2003) 53 *Duke Law Journal* 1, 32–36. It is not only rules that proscribe conduct that can be conduct rules—rules that permit conduct that is otherwise proscribed may also be addressed to, and attempt to guide the conduct of, the general public. So the rule that one may use reasonable force in self-defence is also, at least partly, a conduct rule. I will discuss permissive conduct rules in greater detail in section 2.1 in Ch 2.

³⁴ Gardner, "Decision Rules" 5–6. See also Robinson, 'Rules of Conduct' 732.

as well as to apply subjective criteria inviting individualised discretion and normative assessments.³⁵ This is why decision rules are, and ought to be, framed in relatively indeterminate terms. Decision rules do not aim to provide notice; their primary rationale is to define the rules by which cases can be decided.

It is immediately apparent that conduct rules find a natural home in the *ex ante* stage of the criminal law, which deals with giving notice of rules meant to prevent harmful conduct.³⁶ Offence definitions may be viewed as being primarily³⁷ conduct rules, giving notice of conduct (in the form of *actus reus* and *mens rea*³⁸ stipulations) that should be avoided. They are therefore grounded in objective criteria and state rules of prospective guidance for all persons governed by the criminal justice system. Our experience of crime definition rules tells us that they are usually stated as rules for immediate application, as opposed to rules the application of which is contingent upon the existence of specified circumstances, and that they are usually impersonal, in that they are addressed to all people generally, without reference to the particular characteristics or circumstances of the individual addressees of the norm.³⁹ Furthermore, since they are principally concerned with giving prior notice to the general public in order to prevent harm, they are not primarily concerned with making personal blaming judgments in respect of a person violating the rule.

It is also apparent that the adjudicatory stage of criminal justice is populated by decision rules.⁴⁰ A violation of a general criminal law conduct norm which results in the proscribed outcome (that is, the commission of a *prima facie* offence) activates the jurisdiction of the criminal justice system to punish.⁴¹ The system's

³⁵ Gardner, "Decision Rules" 6.

³⁶ H Stewart, 'The Role of Reasonableness in Self-Defence' (2003) 16 *Canadian Journal of Law and Jurisprudence* 317, 333–34. See also M Thorburn, 'Justifications, Powers, and Authority' (2008) 117 *Yale Law Journal* 1070, 1095.

³⁷ Primarily, but not solely. Offence definitions also function as one of the decision rules in play when determining the blameworthiness of the actor after she has committed a *prima facie* offence.

³⁸ The appropriateness of including a *mens rea* stipulation in a conduct rule has been debated, with GP Fletcher, *Rethinking Criminal Law* (Boston, Little Brown & Company, 1978) 475–79, 553–54, 694–97 arguing that it is essential; Robinson, *Structure and Function* 129–37 arguing that it is mostly unnecessary; and Duff, 'Rule-Violations and Wrongdoings' 68–74 arguing that both views have their flaws. Although I cannot enter into this debate here, I prefer and adopt Fletcher's views, which appear to have gained the most currency.

³⁹ These characteristics, while generally true of offence definitions, need not be true of all conduct rules, or even all conduct rules within the criminal law.

⁴⁰ For instance, Dan-Cohen, 'Decision Rules and Conduct Rules' 637–38 argues that duress and some instances of necessity are decision rules that would, in a world of acoustic separation, be conveyed only to officials. I will argue in this book that other defences too are best viewed as decision rules. See also Stewart, 'The Role of Reasonableness' 10.

⁴¹ By 'activates', I mean that it is only once it is established that a *prima facie* offence was committed that the judge has the jurisdiction to hold the defendant guilty of committing a crime and, thereupon, to punish her. This jurisdiction may be exercised fully, partially or not at all. See M Gur-Arye, 'Justifying the Distinction between Justifications and Power (Justifications vs. Power)' (2010) 5 *Criminal Law and Philosophy* 293, 296, who explains that this is in line with the German criminal law's structure of criminal offences, as per which: 'Conduct that does not satisfy all elements of the definition of the offense remains outside the scope of the criminal law. Conduct that satisfies the elements of the definition

officials must now evaluate, by reference to decision rules addressed to them, the *prima facie* offender's conduct to determine whether she deserves a blaming judgment and, if so, whether she also deserves punishment. These decision rules include rules relating to the appreciation of evidence, the procedure to apply at trial, the appropriate sentence to pass and—critically for the purposes of this study—the availability of defences.

This is not to suggest either that offence definitions are solely conduct rules or that defences are necessarily solely decision rules. As Dan-Cohen recognises, the same rule can be both a conduct rule and a decision rule. However, I do make a more limited assertion here—in an account of the criminal law that is sensitive to the temporal stages in which it functions, the characteristic feature of an offence definition is its nature as a conduct rule, and the characteristic feature of a defence, *qua* defence, is its nature as a decision rule. I will rely upon this stipulation about the characteristic features of offence definitions and defences as I attempt to derive and describe the nature of supervening defences.

1.4 From Supervening to Rationale-Based Defences

The rejection of the wrongness hypothesis has important implications for how we conceive of supervening defences. Consider the following alternative statements made by a defendant who admits to having committed the *actus reus* of an offence with the stipulated *mens rea*:

1. 'I am entitled to a defence because, even though this formed no part of my reasons for acting, my actions did more good than harm.'
2. 'I am entitled to a defence because, even though this formed no part of my reasons for acting, my victim happened to be in the process of attacking me with lethal force and my actions had the effect of repelling the attack.'
3. 'I am entitled to a defence because of my reasons for having acted.'

In principle, all three statements are claims that the defendant is entitled to a supervening defence. However, not all of these claims are compatible with the rejection of the wrongness hypothesis. Statements 1 and 2 above assert (for different reasons) that although the *prima facie* offence was committed, in fact, no wrong (depending on how one defines a wrong) meriting criminal law intervention occurred. But the rejection of the wrongness hypothesis as the basis of supervening defences means that such assertions, even if they are true, cannot

constitutes a criminal wrongdoing which may be negated by justifications.' Hence, the acceptance of a justification is a (negative) exercise of the jurisdiction to blame and punish rather than an acknowledgement that no such jurisdiction ever arose.

themselves ground a supervening defence. In other words, no supervening defence can operate solely by demonstrating that in fact there was no wrong meriting criminal law intervention. Statement 3 is wide enough to include both cases in which on the whole no wrong occurred and cases in which a wrong did occur. Its exculpatory force derives from the defendant's reasons for acting. Only statement 3 can ground a supervening defence that is compatible with the rejection of the wrongness hypothesis.

Another way of looking at this is that at the time that the defendant makes a claim to it, a supervening defence functions purely as a decision rule. At this time, the harm proscribed by the conduct rules has already come to pass and we are now at the *ex post* stage of blame evaluation. To make the argument at this stage that in the instant case, the harm done is not one that merits criminalisation (as opposed to criminal sanction) is as out of place as an argument that raises as a defence to a charge of theft the proposition that theft should not be a criminal offence. Such an argument is an appeal for individual and retroactive legal reform rather a claim to a defence.

The rejection of the wrongness hypothesis then reduces all supervening defences to rationale-based defences, and I think that this is a morally defensible reduction in the scope of the defences available.⁴² Indeed, for many theorists, including those who subscribe to the wrongness hypothesis, a supervening defence is only available if the defendant acts for the right reasons.⁴³ Accordingly, I will henceforth use the term 'rationale-based defence' instead of 'supervening defence', except where the latter is more appropriate for the context.

1.5 A Rationale-Based Alternative to the Wrongness Hypothesis

The rejection of the wrongness hypothesis for justifications means that the standard account of excuses being based on factors that negate the culpability of the agent no longer convinces, since it can no longer explain the distinction between justifications and excuse. Excuses do negate the culpability of the agent, but then so do justifications. What then sets justifications apart from rationale-based excuses?

My hypothesis builds on Karl Binding's suggestion that the normative system that characterises and underlies core rules of the criminal law operates independently

⁴² I will qualify this assertion somewhat in section 3.2.1 in Ch 3.

⁴³ See, for instance, Greenawalt, 'Distinguishing Justifications from Excuses' 95; Fletcher, 'The Right Deed' 320; J Gardner, 'The Gist of Excuses' (1998) 1 *Buffalo Criminal Law Review* 575, 588; AP Simester, 'Wrongs and Reasons' (2009) 72 *MLR* 648, 659, 664.

of the doctrinal law designed by different jurisdictions to capture its content.⁴⁴ I will describe my understanding of the morally derived system of norms that I argue underlies the criminal law⁴⁵ and will then propose that an agent who violates an offence stipulation despite acting for reasons that comport with the criminal law's morally derived normative foundations acts with justification.⁴⁶ In fact, I will argue that this is the paradigmatic form of a justification. However, situations may still arise in which although the normative guidance underlying the criminal law requires that a person act in a particular manner, it would be hypocritical for society to demand this of the person, given that it would not demand the same of itself. In such cases, I propose that the defence made available to the defendant should be classified as an excuse.⁴⁷ In paradigm cases, the grant of an excuse records our judgment that the defendant's reasoning as revealed by her conduct, though morally lacking, conformed to standards that we as a society normatively expect from ourselves when performing the same role as her and that therefore it would be hypocritical for us to single her out for the label of 'evil' or 'criminal'.

My hypothesis, then, is that we can distinguish between justifications and rationale-based excuses on the basis of the quality of the reasoning displayed by the defendant. I therefore refer to it as the 'quality of reasoning hypothesis'. In order to make good my hypothesis, I must demonstrate the following:

1. the manner in which conduct rules and decision rules operate in the criminal law;
2. that the normative system that characterises the core of the criminal law derives from moral norms;
3. the actual structure of the normative system that characterises the core of the criminal law; and
4. the implications of this structure of underlying norms for the contours of justifications and rationale-based excuses, as set up in terms of my hypothesis.

I will attempt to do so in the subsequent chapters of this study.

Only thereafter will I be able to explore in greater detail the manner in which supervening defences ought to work in practice, and the corollaries and limitations generated by the hypothesised view of justifications and defences. The success of the argument will depend in part upon the intuitive plausibility of the entire framework of supervening defences that it generates and in part upon the appeal of the intuitions on which it is founded.

⁴⁴ Because Binding's argument was made in German (in his treatise *Die Normen und ihre Ubertretung*), I rely on Eser's summary of it in Eser, 'Justification and Excuse' 625. Binding separates the notion of 'unlawfulness' from the statutory text intended to capture it, thereby giving unlawfulness an autonomous function. He argues that what is truly violated by a criminal act is not the penal provision which provides a sanction, but instead the commands and prohibitions of the legal system, ie, its norms, which in theory are logically prior to the written law. Accordingly, the concept of wrongfulness must be oriented to and conceived out of the substance of the norms.

⁴⁵ See Ch 3.

⁴⁶ See Chs 4 and 5.

⁴⁷ See Ch 6.