

* *This set of posts is written with a view to being accessible to a wide range of readership. I've therefore kept footnotes and references to a minimum. The ideas I discuss here are set out in much greater detail, with references, bibliography, index, and other bells and whistles in my 2017 monograph, 'Rationale-Based Defences in Criminal Law', which you can find in larger public and university libraries, in addition to being available to purchase online and in bookstores. If you have no way of accessing a copy but would like to check my sources on any of these ideas, please reach out to me at m.dsouza@ucl.ac.uk, and I'll try to help.*

On Justifications and Excuses

Part 1: Could the 'wrongness hypothesis' be wrong?

In this post, I suggest that there is reason to doubt one of the foundational propositions that is often used to structure our understanding and classification of criminal law defences – what I call the 'wrongness hypothesis'. The wrongness hypothesis is used in most theories of criminal defences to separate out the set of rationale-based defences into justifications and rationale-based excuses.

Rationale-based defences

'Rationale-based defences' are defences in which the defendant admits to being a morally responsible agent (so, a sane adult exercising volitional control over their actions), who performed the actus reus of an offence, with its mens rea, but offers up her *reasons for having chosen* to commit the prima facie offence as grounds for exculpation. They include pleas of self-defence, duress, necessity, and certain statutory defences, but exclude defensive pleas like insanity, infancy, automatism, diplomatic immunity, limitation, alibi, and mistake/ignorance of key facts. In this set of posts I focus only on rationale-based defences.

The wrongness hypothesis

The wrongness hypothesis is a widely held view about the difference between justifications (which almost everyone agrees are always rationale-based) and the subset of all excuses that is rationale-based. According to this view, a necessary (though not always sufficient) feature of justifications is that, while admitting the technical actus reus, they negate the objective *all-things-considered* wrongness of the event that happened. Rationale-based excuses on the other hand, leave the objective wrongness of the event and the agent's (D's) mens rea intact, but block the inference of *blameworthiness* for bringing about this objectively wrongful event that would normally follow from these findings, for reasons flowing usually from D's personal circumstances. Together, these

propositions constitute the wrongness hypothesis.

Theories premised on the wrongness hypothesis insist that a legally justified act does not occasion a legal wrong, and an actor who in fact authors a legal wrong cannot be justified in law. Better theories also insist that in order to be justified, the actor should have acted for justified reasons. For such theories, the labels ‘justified’ and ‘excused’ communicate (to the public and to the defendant concerned) a composite judgement about the action and the actor. On this view, a justification says that objectively the right thing happened, *and* that D behaved appropriately in doing what they (subjectively) recognised to be the right thing. A rationale-based excuse says that objectively the right thing did not happen, but that for certain reasons (specified differently by different theories), D should not be blamed for choosing to bring about the event.

Most wrongness hypothesis-based theories, fairly uncontroversially, treat self-defence as a justification, and duress as an excuse. But these theories have more trouble with other defences. There is some disagreement about whether lesser-evils necessity – that form of necessity which involves averting an evil by causing a lesser evil – is a justification or an excuse. Of course, lesser-evils necessity may just be a particularly hard case. But these theories also reach conclusions about cases of mistaken self-defence that we might find unsatisfactory. Consider this example:

Putative Justification: D honestly thinks that V is attacking her with deadly force. Therefore D defends herself by using moderate force against V. In fact D is mistaken. What’s more, D’s mistake was perfectly reasonable – anyone would have thought V was attacking them with deadly force in the same circumstances.

On these facts, because V was not, in fact, attacking D, nothing in D’s explanation for why she used force on V negates the objective all-things-considered wrongness of the harm that V suffered. Therefore, on a wrongness hypothesis-based theory of justification and excuse, D would not be able to claim a justificatory defence – at best, she would be excused from criminal liability. But this doesn’t *feel* right – after all, D could not have done better. Nor could anyone else in D’s position. Yet, D gets, at best, an inferior defence.

Now, there are some ways to respond to this objection – for instance, one could dispute the assertion that an excuse (or at least this particular one) is an inferior defence to justificatory self-defence. Or, one could bite the bullet, and accept and defend this implication of wrongness hypothesis-based theories by insisting that the label of justification is unavailable where V suffers an undeserved harm.

But these responses don't settle the matter. They both make assertions about how we should understand the labels 'justification' and 'excuse', and argue back from them. But since the point of the wrongness hypothesis is to tell us how we should understand these same labels, these are question-begging arguments. To be convinced, we'd need independent reason to think that the chosen understandings of the labels 'justification' and 'excuse' are more useful than rival ones.

Over the course of this series of posts, I'm going to defend a different and, in my view, better, way to distinguish between justifications and excuses. In this post though, I'm just going to argue that we should be open to rejecting the wrongness hypothesis since a plausible alternative may be available.

The beginnings of an alternative

The first thing to say is that although a lot of theories are premised on the wrongness hypothesis, there has been little or no systematic attempt to defend it. That is probably because there is no pre-ordained uniquely correct way to use the terms justification and excuse when dividing up rationale-based defences and so there is no knock-down argument for or against using the wrongness hypothesis to shape our usage of these terms. Even so, the point stands – there is no unassailable reason to think that the wrongness hypothesis is correct. If we find a viable alternative that is as good or better, we should be open to accepting that instead.

Which leads to the next point: to compare rival theories about how we should understand justifications and excuses, we should consider the rational appeal of these theories, and the plausibility of the systems they generate. 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 just like 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'. 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, no (non-intuited) reason has been offered to substantiate the assumption that there should be an actus reus/justification and mens rea/excuse symmetry. True, offences and defences (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. Or at least, they do not found the only account of how the criminal law functions.

Another plausible account – 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 D’s blameworthiness. On this account, offence stipulations (which set out the actus reus and mens rea requirements) belong in the *ex ante* stage of criminal law. Therefore, their features should be adapted to the function that they perform, viz. providing prior conduct guidance with a view to avoiding undesirable happenings, such as, the death of a human at the hands of another. In offering prospectively guidance on how to avert that happening, the offence stipulation might specify, for instance, that people should not set out to cause the death of others. But rationale-based defences come into play only *after* the happening, and once D has admitted to acting in contravention of the advance guidance, i.e. after D admittedly chose to kill another and succeeded. These defences are raised in court to protect the defendant against a blaming judgment; it is, after all, the defendant, and not the deed, on trial. Arguably then, their features should relate primarily to the function they perform at this stage, viz. undermining the defendant’s apparent blameworthiness. On this view, *all* rationale-based defences function by negating (at least some aspects of) the defendant’s blameworthiness, and there is no reason to expect any of them necessarily to negate the occurrence of a wrongful outcome. The important question in a criminal trial is whether *the defendant* was justified or excused; not whether *the event* that resulted from the defendant’s intervention was justified or excused.

You might worry that on this account, a person can be fully justified even when a wrongful outcome occurs, and so we have no normative space to ascribe moral value to what actually happened. But that isn’t quite right. The space we need is available at the *ex ante* stage of deciding what should be a crime, and what label it should carry. For instance, if we decided that ‘X’ was a wrong that should be discouraged using the criminal law, we could draft an *ex ante* offence stipulation that would only be violated when X actually happened (rather than when the actor tried to cause X, or did not take care not to cause X). Then, when X fortuitously did not happen despite the actor’s best efforts, the stipulation against causing X would not be violated, and the actor would attract no blame vis-a-vis that stipulation (although she may well attract blame vis-a-vis other stipulations).

Still you might think, this account must suggest some way of distinguishing justifications from excuses to *really* compete with the wrongness hypothesis. That’s what this next section starts to develop.

Conduct rules and decision rules

I mentioned that the criminal law works in stages, with an *ex ante*, forward-looking stage, aimed at preventing undesirable occurrences by telling us in advance what conduct to avoid; and an *ex post*, backward-looking stage, in which it responds to apparent violations of the *ex ante* guidance by people who have committed a *prima facie* offence. Here's another way to think of this idea: we can categorise the rules in the criminal law system into 'conduct rules' and 'decision rules'. Conduct rules are aimed at preventing bad things from happening by addressing the general public and setting out advance guidance about conduct to be avoided on pain of criminal sanction. Decision rules aren't concerned with preventing specific bad things from happening (except insofar as we accept some theory of general deterrence) – they come into play only after a seemingly bad thing has already happened. They are addressed to officials like the police, prosecutors, and judges and guide them on how to perform their roles in response to this seemingly bad thing. Because the primary rationale of conduct rules is to give notice, they should be drafted to be intelligible to, and usable by, the public. Accordingly, they should be simple, brief, and call for as little exercise of judgement (which brings with it uncertainty) as possible. Decision rules, on the other hand, aim to provide guidance only to people empowered and specially trained to exercise their judgement – the officials that administer the criminal justice system. Their training helps these officials to evaluate the complex and varied situational factors of particular cases, and exercise their judgement in reaching decisions. Accordingly, these rules can be, and often are, framed in relatively broad and open-ended terms that call for the exercise of judgement. Any given rule may be a conduct rule, or a decision rule, or both.

From our experience of offences and defences, it is immediately apparent that offence definitions are primarily conduct rules, and rationale-based defences are primarily decision rules. Yet offence definitions are also partly decision rules, insofar as officials will need to make their determinations also by reference to them, and some defences – specifically, justifications – are also conduct rules, since they identify conduct that is not merely reprieved from punishment, but actually legally permissible (and sometimes, encouraged). This last observation allows us to distinguish justifications from rationale-based excuses. Justifications are primarily decision rules though they also offer some conduct guidance. Excuses are decision rules that are not generally thought to provide any conduct guidance whatsoever. Put another way, we can distinguish between justifications and excuses based on the quality of the defendant's exculpatory reasoning. Justifications exculpate when the defendant's reasons for having performed a *prima facie* offence conform to the conduct guidance contained in the criminal law system (even though this conduct guidance is not captured by the offence stipulation). Excuses exculpate even though the defendant's reasons for having committed a *prima facie* offence do not conform to the normative system of conduct guidance in the criminal law system. I call this

competitor to the wrongness hypothesis, the ‘quality of reasoning’ hypothesis.

I need to say a lot more to flesh out the quality of reasoning hypothesis. A fuller account of it would

- (a) explain how conduct and decision rules map onto doctrinal criminal law;
- (b) describe the criminal law’s normative system of conduct guidance;
- (c) expand on the source of exculpation in excuses;
- (d) state and defend a plausible view about whether lesser-evils necessity is a justification or an excuse or something else altogether; and
- (e) detail the implications of all these propositions for the contours of rationale-based defences.

I will address these matters in subsequent posts, before finally working through the corollaries and limitations generated by the hypothesised view of justifications and excuses. The success of my argument will depend in part upon the intuitive plausibility of the entire framework of rationale-based defences that it generates, and in part upon the appeal of the intuitions on which it is founded.

** This post summarises the argument in Chapter 1 of my 2017 monograph, ‘Rationale-Based Defences in Criminal Law’. For much more detail, please consult that chapter. I recommend reading the Introduction to the monograph as well – this sets out the overall plan of action for the monograph, and explains how I hope to convince readers about my thesis despite it addressing a puzzle in criminal law theory to which there is no uniquely correct answer.*