

ISSN 1936-5349 (print)
ISSN 1936-5357 (online)

HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

EQUITABLE DEFENCES AS META-LAW

Henry E. Smith

Published in *Defences in Equity*
Paul S. Davies, Simon Douglas & James Goudkamp, eds., Hart (2018)

Discussion Paper No. 954

03/2018

Harvard Law School
Cambridge, MA 02138

This paper can be downloaded without charge from:

The Harvard John M. Olin Discussion Paper Series:
http://www.law.harvard.edu/programs/olin_center

The Social Science Research Network Electronic Paper Collection:
<http://ssrn.com/abstract=3134804>

Equitable Defences as Meta-Law

Henry E Smith* (forthcoming in *Defences in Equity*,
Paul S. Davies, Simon Douglas & James Goudkamp, eds., Hart 2018)

Abstract: Equitable defences are the front line of controversy over fusion. Because law and equity offer a range of defences that partially overlap and the rationale for matching equitable defences to equitable remedies is at least as obscure as the rationale for separate equitable remedies, conventional wisdom holds that the more one can fuse the equitable defences into the law the better. In this chapter I argue that equity roughly reflects a distinct function – a safety valve that operates at a higher (meta) level over the rest of the law and that responds to problems of high uncertainty and variability. These characteristic problems include opportunism and multipolar conflicts. From a functional point of view, some of the special treatment of equitable defences makes sense, and puzzling patterns in this area receive an explanation and some justification. Even the adaptation of some equitable defences into the law dovetails with a dynamic picture of the equitable function.

I. Introduction

Is there anything special about equitable defences? Historically, equity courts developed defences that displayed certain characteristics. Equitable defences tended to be more discretionary, to sound

* Fessenden Professor of Law and Director of the Project on the Foundations of Private Law, Harvard Law School. For helpful questions and comments, I would like to thank participants at the Workshop on Defences in Equity, Jesus College, Oxford, and a faculty workshop at the Harvard Law School. Thanks to Molly Brown for excellent research assistance. All errors require a defence on my part.

in morality and to defeat remedies – themselves denominated equitable – rather than cutting off liability altogether. Nevertheless, treating these defences as more than an historical happenstance or institutional artifact is difficult, because common law defences sometimes involve discretion and reflect moral considerations. And saying that an equitable defence is paired with an equitable remedy leaves us with the question of what is uniquely equitable about remedies like an injunction, reformation or accounting. Conversely, many equitable defences do not look all that discretionary or uniquely infused with morality. And after the fusion of law and equity, the pairing of equitable defences with ‘equitable’ remedies has partially broken down, and looks ripe for further assimilation.¹ So it would appear that treating equitable defences as an interesting or justifiably separate category would be a tall order.

In this chapter, I argue that there is an interesting and functionally justified notion of ‘equitable defence’. It is not coextensive with what went under that heading when there were separate equity courts. And yet I will argue that the law-equity divide is not irrelevant to a functional theory of equitable defences. Because equity drew on a tradition in which equity corrects law where law fails owing to its generality and because equity intervened literally from outside the law, functional equity and functionally equitable defences in particular tended (but only tended) to originate in equity jurisdiction.

What is an equitable defence, functionally speaking? In related work I develop a theory of the equitable function, in which equity serves as a second-order safety valve on the regular law.² The ‘regular’ part of the law, precisely because it aspires to be general, faces difficulties with hard-to-foresee problems that disturb the stable relationships between activities. Equity on this account

¹ See, eg, A Burrows, ‘We Do This at Common Law But That in Equity’ (2002) 22 *Oxford Journal of Legal Studies* 1; D Laycock, ‘The Triumph of Equity’ (Summer 1993) 56 *Law and Contemporary Problems* 53; A Mason, ‘Fusion’ in S Degeling and J Edelman (eds), *Equity in Commercial Law* (Sydney, Lawbook Co, 2005). See also A Schwartz, ‘The Case for Specific Performance’ (1979) 89 *Yale Law Journal* 271, 298–303.

² HE Smith, ‘Why Fiduciary Law Is Equitable’ in AS Gold and PB Miller (eds), *Philosophical Foundations of Fiduciary Law* (Oxford, Oxford University Press, 2014) 261–84; HE Smith, ‘Fusing the Equitable Function in Private Law’ in K Barker, K Fairweather, and R Grantham (eds), *Private Law in the 21st Century* (Oxford, Hart Publishing, 2017); HE Smith, ‘Equity as Second-Order Law: The Problem of Opportunism’ (unpublished manuscript, 15 January 2015) available at <http://ssrn.com/abstract=2617413>.

is a function that consists of second-order intervention into the rest of the law. That is, equity is law about law; equity refers to law but not vice versa. Some defences serve this equitable function, and for that reason deserve the name ‘equitable’.

This chapter will begin in Section II by setting out how some law can function as meta-law and why historically equity courts focused on serving this function. Section III turns to special considerations that arise when defences, many of them historically equitable, serve a second-order function in solving problems of uncertainty and variability, especially party opportunism. Section IV explores the dynamic dimension of equity and shows how the equitable function leads to new law. The chapter ends with some thoughts on the implications of equitable defences as meta-law for the debates over the fusion of law and equity.

II. Equity Functioning as Meta-Law

Many aspects of law are second or higher order – they operate at a meta level. For example, a power is second order in the sense that one who has a power can change legal relations: the power refers to (and potentially changes) these legal relations. But not vice versa: those lower-order legal relations do not make reference to the power that might change them.³ Thus, a power to transfer is the power to change who holds a right (and who bears the corresponding duties). A power to legislate is the power to create law. Even courts in the common law when adjusting legal rules to fit new conditions, in a sort of quasi-legislation, are exercising a power – a power to change the law.⁴ They are engaged in a ‘meta’ enterprise: the judicial law-creating process makes reference

³ See WN Hohfeld, ‘Fundamental Legal Conceptions as Applied in Judicial Reasoning’ (1913) 23 *Yale Law Journal* 16, 44–58; U Pagano, ‘Public Markets, Private Orderings and Public Governance’ (2000) 20 *International Review of Law and Economics* 453, 459–65; TM Sichelman, ‘Quantifying Legal Entropy’ (2013) San Diego Legal Studies Paper No 13-128, available at <http://ssrn.com/abstract=2293015>. Although I do not rule out that there may be higher orders than second, for present purposes I will deal with questions of meta-law as involving a second order.

⁴ It is interesting that the traditional aversion to admitting this casts this function as sounding as if it were not second order at all.

to the law, which by its terms need not make reference to this (higher-order) process. Thus, there are many aspects of law and the legal system that could be termed ‘meta’.⁵

I argue that there is a special kind of meta-law that deals with problems that become too complex and unforeseeable to be easily handled at the primary level on which they arise. The most compelling and immediate problem faced by a system of ‘general’ law is efforts by actors within the system of finding its weak points and exploiting them. This is a familiar problem in tax law but is true across law.⁶ For example, what happens when someone uses the possibility of an injunction solely to extract payment from a rights violator out of all proportion to the actual injury suffered? What if someone invokes the letter of a contract in order to evade payment for the full value received, as where a builder inadvertently uses a pipe that is the wrong brand but that is of equal quality?⁷ One could always imagine the law being better spelled out, contracts covering more contingencies, or some source supplying better and more complete information to primary actors, but none of this is costless. The fear is that as soon as these holes are plugged, opportunists will

⁵ The ‘rules about rules’ in Robert Stevens’ chapter in this volume (Chapter 3) are meta, and it worth pointing out that equitable set-off involves adjustment to primary claims in the light of relationships between claims that are variable, complex and hard to foresee. Also ‘meta’, in a sense, are equitable rights if one views them as rights against rights. See B McFarlane and R Stevens, ‘The Nature of Equitable Property’ (2010) 4 *Journal of Equity* 1.

⁶ DA Weisbach, ‘Formalism in the Tax Law’ (1999) 66 *University of Chicago Law Review* 860. See also SB Lawsky, ‘Probably? Understanding Tax Law’s Uncertainty’ (2009) 157 *University of Pennsylvania Law Review* 1017, 1032, arguing that tax law uses probabilistic doctrines because ‘the essence of a tax shelter is that it technically complies with the law while nonetheless violating the substance or intent of the law, which is no easy thing to determine’; SS Surrey, ‘Complexity and the Internal Revenue Code: The Problem of the Management of Tax Detail’ (1969) 34 *Law and Contemporary Problems* 673, 707, fn 31.

⁷ *Jacob & Youngs v Kent* 129 NE 889 (NY 1921); VP Goldberg, ‘Rethinking *Jacob & Youngs v. Kent*’ (2015) 66 *Case Western Reserve Law Review* 111; HE Smith, ‘Is Equitable Contract Law a Pipe Dream?’ (*New Private Law*, 9 June 2016), available at <http://blogs.harvard.edu/nplblog/2016/06/09/is-equitable-contract-law-a-pipe-dream-henry-smith/>.

find yet others to exploit.⁸ One option is to do the best one can at the primary level and then tolerate whatever bad behaviour falls through the cracks.

There is yet another solution: equity. By ‘equity’ I mean a part of the law serving a particular function.⁹ For historical and institutional reasons this function was a theme – but only one theme and not the exclusive preserve of equity courts.

What is the equitable function and how is it served? In systems theory, problems of high variance and uncertainty sometimes call for a solution at a higher level.¹⁰ In hierarchy, the higher-order component makes reference to the primary component but not vice versa.¹¹ Think of part of a computer program that keeps values at a primary level within a certain range. Or a safety valve that kicks in when a threshold is reached. I argue that equity serves a function much like this in law.

Equity is a second-order safety valve on the law, which responds to problems of special complexity and unforeseeability. These include not only opportunism, but also conflicting rights and multipolar interactions.¹² For example, certain problems in nuisance require an equitable style of analysis in which the interaction between two potential uses is analysed in terms of what

⁸ This is one reason why parties may contract in anticipation of some opportunism, but the law will not allow parties to bargain out of the duty of good faith. Opportunism may also arise in the context of litigation over remedies. See *Ruxley Electronics and Construction Ltd v Forsyth* [1996] AC 344 (HL), in which the judge took the plaintiff’s offer to use the damages to rebuild a swimming pool, which did not conform to the contract but was equivalent in market price, as vindictive and extortionate. See also Smith (n 7).

⁹ See the sources cited above at n 2.

¹⁰ See, eg, AY Aulin-Ahmavaara, ‘The Law of Requisite Hierarchy’ (1979) 8 *Kybernetes* 259; F Heylighen and C Joslyn, ‘Cybernetics and Second-Order Cybernetics’ in RA Meyers (ed), *Encyclopedia of Physical Science and Technology* 3rd edn (San Diego, Academic Press, 2002).

¹¹ See, eg, KL von Bertalanffy, *General System Theory* (New York, Braziller, 1968). See also JH Holland, *Hidden Order* (Reading MA, Addison-Wesley, 1995) 11–12, discussing second-order agents and properties.

¹² Smith, ‘Fusing the Equitable Function in Private Law’ (n 2); Smith, ‘Equity as Second-Order Law’ (n 2).

maximises mutual freedom or utility, in a symmetric fashion.¹³ Multipolar problems, which Lon Fuller, borrowing from Michael Polanyi, called ‘polycentric’, involve multiple parties and many interdependencies. Fuller gave as an example the problem of dividing up a collection of paintings left under a will to two museums without instructions as to which paintings would go to which museum.¹⁴ The problem is complex because the value to a museum of each painting depends on the other paintings it gets. The interdependencies here make this a problem like those familiar from complexity theory, in which the time required to solve the problem sometimes increases exponentially in the size of the problem.¹⁵ In complexity theory, such problems tend to call for approximate solutions.

They also often call for solutions at a higher level. Consider opportunism. Anticipating opportunism is inherently difficult. By evaluating it *ex post*, the court or other decision-maker has the advantage of moving second. The opportunist in some sense operates at a higher level than other parties: such actors have the entire system in view and exploit it accordingly. By moving to the same higher level, courts and other decision-makers can occupy the same ground. *Ex post* and somewhat context-dependent decision-making also can be tailored to the problem presented by a specific instance of opportunism.

¹³ JCP Goldberg and HE Smith, ‘Wrongful Fusion: Equity and Tort’ in JCP Goldberg, HE Smith and PG Turner (eds), *Equity and Law: Fusion and Fission* (Cambridge, Cambridge University Press, forthcoming). A limited equitable function within and around the law of tort does not amount to general equitable jurisdiction over torts, much less crimes.

¹⁴ LL Fuller, ‘The Forms and Limits of Adjudication’ in KI Winston (ed), *The Principles of Social Order: Selected Essays of Lon L. Fuller* (Oxford, Hart Publishing, 2001) 126–36, introducing the concept of polycentric tasks. See also M Polanyi, *The Logic of Liberty; Reflections and Rejoinders* (Chicago, University of Chicago Press, 1965) 171.

¹⁵ A famous similar problem is the so-called ‘knapsack problem’, in which one is required to choose the combination out of a given set of n items that will maximise value under a given weight limit. This problem requires exponential time (as n increases) and is in the class of probably intractable problems known as ‘NP-complete’. See R Greenlaw and HJ Hoover, *Fundamentals of the Theory of Computation: Principles and Practice* (San Francisco, Morgan Kaufmann, 1998) 287–313; K Devlin, *The Millennium Problems: The Seven Greatest Unsolved Mathematical Puzzles of Our Time* (New York, Basic Books, 2002) 105–30.

Unconstrained, such a second-order intervention is a powerful tool, and it is correspondingly destabilising. If it could in principle have any effect based on any combination of circumstances at the primary level, it would create a great deal of uncertainty in its results and would undermine the guidance function of the law. Opponents of equity were quick to point out this tendency, and accused equity courts of going too far down this road. Hence the gibe about the ‘Chancellor’s Foot’.¹⁶

As with second-order interventions generally, the conditions for going to the second level can prevent this unravelling. In a safety valve, there are conditions that delimit when it becomes relevant. So too in equity, the basic presumption is for law, and it can only be overcome when some triggering conditions are met. Some of these triggers are proxies for opportunism, which in turn operate in a domain of potential application.

In traditional equity, then, the domain of potential applicability of equity was the familiar trio of ‘fraud, accident, and mistake’.¹⁷ Crucially, equity has no roving commission to root out

¹⁶ The most famous critique of equity is that of John Selden, ‘Equity’ in R Milward (ed), *Table-Talk: Being the Discourses of John Selden, Esq* (London, JM Dent & Co, 1898) 43, 43–44: ‘Equity is a Roguish thing: for law we have a measure, know what to trust to; Equity is according to the Conscience of him that is Chancellor, and as that is larger or narrower, so is Equity. ‘Tis all one as if they should make the Standard for the measure we call a Foot, a Chancellor’s Foot; what an uncertain Measure would be this. One Chancellor has a long Foot, another a short Foot, a Third an indifferent Foot: ‘Tis the same thing in the Chancellor’s Conscience’. See generally JH Baker, *An Introduction to English Legal History* 4th edn (London, Butterworths LexisNexis, 2002) 96–115.

¹⁷ *American Jurisprudence* 2nd edn (2016) vol 47, ‘Judgments’ § 718, Westlaw (database updated November 2015): ‘Generally, claimants seeking equitable relief from judgments through independent actions must meet three requirements [the third of which is that] they must establish a recognized ground, such as fraud, accident, or mistake, for the equitable relief’ (footnotes omitted, citing cases). See also WF Walsh, *A Treatise on Mortgages* (Chicago, Callaghan and Co, 1934) 6, 11, regarding relief from mortgages in equity on grounds of fraud, accident or mistake; VD Ricks, ‘American Mutual Mistake: Half-Civilian Mongrel, Consideration Reincarnate’ (1998) 58 *Louisiana Law Review* 663, 717, and see also at fn 277, speculating that Chief Justice Allen in *Swift v Hawkins* 1 Dall 17 (Pa 1768) ‘considered “mistake” to be representative of all categories of equity’. A poetic version is attributed to Thomas More, the first lawyer to serve as Chancellor: ‘Three things are to be helpt in Conscience; Fraud, Accident and things of Confidence’, 1 Rolle’s Abridgement 374. See also *Coco v AN Clark (Engineers) Ltd* [1969] RPC 41, 46 (Ch) (Megarry J), quoting More’s couplet; A Laussat Jr, *An Essay on Equity in Pennsylvania* (Philadelphia, Robert Desilver, 1826)

opportunism wherever it can be found. On the contrary, equity only kicks in when it is triggered. The triggers for equity as an anti-opportunism device sound in bad faith and disproportionate hardship. The exact triggers differ somewhat depending on the class of problems involved. Thus in building encroachments, a continuing trespass gives rise to a presumption for an injunction. But if the injunction would visit much greater hardship on the defendant than on the plaintiff, a court may withhold the injunction, leaving the plaintiff to a damages remedy.¹⁸ Nevertheless if the encroachment was in bad faith, which in the context of building means knowledge of the boundary, the defence of disproportionate (or undue) hardship would be unavailable and the injunction would issue. In contracts, the unconscionability defence works similarly, the trigger being a combination of a startlingly one-sided result and a vulnerable party. Once this trigger has been satisfied, the contract is subjected to some version of a closer scrutiny on fairness grounds, an analysis in which conscious exploitation plays an important role.

Much of the danger in fraud, accident, and mistake comes from the opening they afford for advantage-taking. In traditional parlance, the problem was ‘constructive fraud’, which not atypically takes up the bulk of Justice Story’s treatise. He describes constructive fraud:¹⁹

There is always fraud presumed or inferred from the circumstances or conditions of the parties contracting, weakness on one side, usury on the other, or extortion or advantage taken of that

67, stating that ‘Sir Thomas More used to say that the following doggerel contained all the heads of chancery jurisdiction’.

¹⁸ *American Jurisprudence* 2nd edn (2005) vol 42, ‘Injunctions’ § 35: ‘Even if the wrongful acts are indisputable, an injunction may be denied if the payment of money would afford substantial redress and if the injunction would subject the defendant to grossly disproportionate hardship’; D Laycock, ‘The Neglected Defense of Undue Hardship (and the Doctrinal Train Wreck in *Boomer v. Atlantic Cement*)’ (2011) 4 *Journal of Tort Law*, Issue 3, Article 3.

¹⁹ J Story, *Commentaries on Equity Jurisprudence, As Administered in England and America* (Boston MA, Hilliard, Gray & Co, 1836) § 334. See also *Nocton v Ashburton* [1914] AC 932, 954 (Viscount Haldane LC): ‘What it really means in this connection [“constructive fraud”] is, not moral fraud in the ordinary sense, but breach of the sort of obligation which is enforced by a Court which from the beginning regarded itself as a Court of conscience’; JN Pomeroy, *A Treatise on Equity Jurisprudence* (San Francisco, Bancroft-Whitney Co, 1907) §§ 922–23. One modern approach is to sever the connection with fraud by replacing the notion of ‘constructive fraud’ with undue influence and unconscionable transactions. See John McGhee (ed), *Snell’s Equity* 33rd edn (London, Sweet and Maxwell, 2015) ch 8.

weakness. There has always been an appearance of fraud from the nature of the bargain, even if there be no proof of any circumvention, but merely from the intrinsic unconscionableness of the bargain.

The notion of constructive fraud is how the law captures opportunism. Whether opportunism is a useful category at all has been controversial in economics for reasons analogous to the debate about equity. Elsewhere I have defined opportunism as ‘behavior that is undesirable but that cannot be cost-effectively captured – defined, detected, and deterred – by explicit ex ante rulemaking ... It often consists of behavior that is technically legal but is done with a view to securing unintended benefits from the system, and these benefits are usually smaller than the costs they impose on others’.²⁰ The nature of opportunism often calls for second-order treatment, which will inevitably be somewhat ex post and open-ended.²¹ As Story put it his treatise, equity had to be open-ended and flexible because ‘[f]raud is infinite’ given the ‘fertility of man’s invention’.²²

The debate among economists about opportunism offers instructive parallels. The most famous invocation of opportunism can be found in the work of Nobel laureate Oliver Williamson, who defines opportunism as ‘self-interest seeking with guile’.²³ Apparently Williamson sees any rule-breaking as opportunism. And yet because not all rule-breaking causes the kind of complexity and uncertainty at issue here, not all rule-breaking calls for extraordinary intervention, much less at a meta level. It is Williamson’s ‘guile’ that seems special. Likewise in other definitions of opportunism based on defeating legitimate expectations and the like, we see undistilled hints of what calls for meta level treatment.²⁴

²⁰ Smith, ‘Equity as Second-Order Law’ (n 2) 14–15.

²¹ Smith, ‘Fusing the Equitable Function in Private Law’ (n 2); Smith, ‘Equity as Second-Order Law’ (n 2).

²² Story (n 19) § 186 (at 196) fn 4, quoting a letter from Lord Hardwicke to Lord Kames (June 30 1759). Chancellor Ellesmere made a similar point in *The Earl of Oxford’s Case* (1615) 1 Ch Rep 1, 6; 21 ER 485, 486: ‘The Cause why there is a Chancery is, for that Mens Actions are so divers and infinite, That it is impossible to make any general Law which may aptly meet with every particular Act, and not fail in some Circumstances’.

²³ OE Williamson, *The Economic Institutions of Capitalism* (New York, Free Press, 1985) 47.

²⁴ See, eg, GM Cohen, ‘The Negligence-Opportunism Tradeoff in Contract Law’ (1992) 20 *Hofstra Law Review* 941, 957, defining ‘opportunism’ as ‘any contractual conduct by one party contrary to the other party’s reasonable expectations based on the parties’ agreement, contractual norms, or conventional morality’ (footnotes omitted); TJ

On the other side are economists who think that opportunism is uninteresting and unimportant. All behaviour is self-interested and any such behaviour within the constraints offered by institutions is legitimate; if not, then the rules of the game need to be changed.²⁵ In a sense, these economists are arguing that behaviour and institutions should be evaluated on one level. Behaviour is behaviour and sets of rules either constrain it cost-effectively or they do not. This view also resonates with a view about rights: that a right is not a right if one cannot exercise it without being second-guessed as to motive.²⁶ For some, abuse of a right is not a legitimate doctrine; if there is a problem with certain exercises of the right, then the only choices are to redefine rights or to tolerate the behaviour in question.

In the present account of equity, the ambition is not to capture it perfectly. Instead, the presumptions and the triggers that overcome them aim to sweep in a mix of behaviour that contains enough opportunism that closer and more open-ended scrutiny will do more good than harm. If this is how the equitable function works it is understandable that it is often mistaken for something

Muris, 'Opportunistic Behavior and the Law of Contracts' (1981) 65 *Minnesota Law Review* 521, 521, defining 'opportunism' as conduct that is 'contrary to the other party's understanding of their contract, but not necessarily contrary to the agreement's explicit terms'. See also, eg, SW Buell, 'Good Faith and Law Evasion' (2011) 58 *UCLA Law Review* 611, 623: 'In common parlance, the evasive actor is one whose project is to get around the law. She seeks to avoid sanction while engaging, in substance, in the very sort of behavior that the law means to price or punish'. For a wider definition, see, eg, RA Posner, *Economic Analysis of Law* 5th edn (New York, Aspen Law & Business, 1998) § 4.1 (at 103), defining 'opportunism' in the contracting context as 'trying to take advantage of the vulnerabilities created by the sequential character of contractual performance'.

²⁵ Y Barzel, 'Transaction Costs: Are They Just Costs?' (1985) 141 *Journal of Institutional and Theoretical Economics* 4, 10–11, arguing that because it is costly to tell whether someone is acting in a self-seeking manner and not 'as he should', the concept of opportunism does not help in understanding behaviour and only confuses; P Klein, 'Does Transaction Cost Economics Need Opportunism?' (*Organizations and Markets*, 6 October 2006), available at <http://organizationsandmarkets.com/2006/10/06/does-transaction-cost-economics-need-opportunism/>. See also OE Williamson, 'Opportunism and Its Critics' (1993) 14 *Managerial and Decision Economics* 97, defending the usefulness of the notion of opportunism against social science critics.

²⁶ See, eg, L Smith, 'The Motive, Not the Deed' in J Getzler (ed), *Rationalizing Property, Equity and Trusts: Essays in Honour of Edward Burn* (London, LexisNexis UK, 2003) 66; N Shoked, 'Two Hundred Years of Spite' (2016) 110 *Northwestern University Law Review* 357, 371.

else. A determinate approach at the primary rather than a meta level would require something very different. Modern commentators such as Douglas Laycock tend to dismiss notions like the irreparable injury ‘rule’ as not constraining because one can point to a wide variety of situations in which it does not prevent a court from ordering specific performance or issuing an injunction.²⁷ These critiques are well-taken if equity were first order, and yet invocations of irreparable injury may make sense as a signal that a court is overcoming the presumption against equity. Indeed, if as Laycock argues, it would be better to replace notions like irreparable injury and equity with functionally justified substitutes, we should consider whether something like the irreparable injury requirement – and equity – can be justified as a method of implementing the two-tier architecture.

Interestingly, the architecture of presumptions and meta levels surfaces occasionally in moral philosophy as well. In threshold deontology a moral principle holds until it crosses some threshold: in an extreme enough situation, other (for example, consequentialist) considerations can come into play. Quite parallel to the idea of an equitable safety valve in law is Philip Pettit’s idea that we should be guided by general dispositions until certain ‘contextual cues’, ‘red lights’ or ‘alarm bells’ go off and then we should be guided by the ‘stand-by guide’ of general principles of right.²⁸ Pettit offers as an example the old joke that a good friend will help you move an apartment but only a very good friend will help you move a body.²⁹ One should be guided by an automatic loyalty to friends until a red light goes off – as when a murder might be involved. Pettit develops a model of moral psychology in which the sensitivity to cues for switching to the stand-by strategy is a skill, which he likens to Aristotle’s ‘phronesis’ and Aquinas’ ‘prudentia’.³⁰ To this we can add

²⁷ D Laycock, *The Death of the Irreparable Injury Rule* (Oxford, Oxford University Press, 1991); D Laycock, ‘The Death of the Irreparable Injury Rule’ (1990) 103 *Harvard Law Review* 687, 688–701.

²⁸ P Pettit, *The Robust Demands of the Good: Ethics with Attachment, Virtue, and Respect* (Oxford, Oxford University Press, 2015) 221.

²⁹ *ibid.*

³⁰ *ibid.* 222. RC Bartlett and SD Collins (trs), *Aristotle’s Nicomachean Ethics* 3rd edn (Chicago, University of Chicago Press, 2011) 120–21 [Book 6 ch 5, 1140a24ff]; St T Aquinas, *Summa Theologiae*, (Fathers of the English Dominican Province trs, Westminster, Christian Classics, 1981) IIaIIae 47.2 ad 1, IIaIIae 47.4.

that the skill also requires knowing how not to switch to the stand-by strategy too quickly. This kind of wisdom helps connect the general and particular.

Equity works similarly. A certain prudence or practical wisdom is necessary for the implementation of equity. Only when certain contextual cues – I have been calling them proxies or triggers – are invoked does one switch from general and simple rules to a more open-ended and direct employment of principles.

Note that the stand-by – or safety-valve – role for equity implements a kind of modularisation and specialisation. The two-tier structure can help to overcome complexity. The second-order safety valve is a specialised module connected by an interface of the proxies and presumptions, the triggers for going from ‘law’ to ‘equity’. Modularity in general helps manage complexity by breaking a complex system into parts and allowing them to interact only in certain ways.³¹ A complex system is one in which it is hard to predict the properties of the whole from the properties of its parts, and the effect of modularity is to contain such ripple effects. A two-tiered or hierarchical structure of modules is a familiar one that is suited for certain kinds of problems involving uncertainty and variability. In functional equity, it is the reservation of some problems, such as opportunism or multipolar problems, to be solved at the second, equitable level that allows first-order rules to be more formal – less contextual and less complex – than they would otherwise be. In other words, functional equity is a module with a largely one-way connection (interface) to the rest of the law. It is triggered by proxies (the stand-by system is conditioned on an alarm) and once triggered it makes use of a wide range of information at the primary level. That primary level does not make reference to the second level. The primary level can be simpler, more general and hence easier to use, if it is backstopped with a resort to general considerations. The triggers (alarm bells) are a simple method for cabining the powerful and complex tool of applying general

³¹ See, eg, CY Baldwin and KB Clark, *Design Rules: The Power of Modularity*, vol 1 (Cambridge MA, MIT Press, 2000); R Garud, A Kumaraswamy and RN Langlois (eds), *Managing in the Modular Age: Architectures, Networks and Organizations* (Malden MA, Blackwell, 2002); HA Simon, *The Sciences of the Artificial* 2nd edn (Cambridge MA, MIT Press, 1981) 195–200; RN Langlois, ‘Modularity in Technology and Organization’ (2002) 49 *Journal of Economic Behavior & Organization* 19; R Sanchez and JT Mahoney, ‘Modularity, Flexibility, and Knowledge Management in Product and Organization Design’ (1996, Winter Special Issue) 17 *Strategic Management Journal* 63.

considerations – freer use of context. Because the freer use of context is confined in this manner, it can be more thoroughgoing and effective than if it were used more generally.

It is probably no accident that equity and moral psychology would share such a structure. A possible reason for their similarity is that they both respond to the same considerations of system organisation. The demands of managing complexity and achieving the goals of ‘correct’ decision-making lead to hierarchy and specialisation through modularity. Using the stand-by system only – all equity all the time – would be complicated and prone to error. And using the general rules would fail spectacularly in extreme circumstances.

Indeed, the idea that equity saves law from bad results in extreme situations resonates with aspects of the equitable tradition. Much of that tradition, including cases and commentary, invokes the Aristotelian idea that equity is ‘a correction of law in the respect in which it is deficient because of its being general’.³² What sets theories of equity apart is what counts as a failure of law owing to its generality. The broadest version of equity would aspire to capture every occurrence of unfairness or every ex post instance of lack of fit between rules and their purposes. On the safety-valve account offered here, equity is not primary. Instead, primary law is meant to guide conduct and to apply generally, not on a case-by-case basis. But the attempt to be general runs up against limits from complexity and lack of foreseeability.

Modularity allows equity to specialise as well in its degree of formalism relative to law. Within the domain of equity, decision-makers are able to make freer use of context than they usually can. It is a truism that equity is less formal than law, but it is true in a deep sense. Formalism can be defined as relative invariance to context.³³ Thus English is less formal than computer languages, and the notation of everyday mathematics is less formal than that of published proofs. In everyday speech, pragmatic meanings like ‘Please close the window’ when actually uttering ‘It’s cold in here’, make colloquial speech very context-dependent. Law strips a lot of context out

³² Bartlett and Collins (n 30) 112 [Aristotle, *Nicomachean Ethics* 1137b, lines 26–28].

³³ F Heylighen, ‘Advantages and Limitations of Formal Expression’ (1999) 4 *Foundations of Science* 25, 25, 37; HE Smith, ‘The Language of Property: Form, Context, and Audience’ (2003) 55 *Stanford Law Review* 1105, 1112–13, 1135–39, 1181–90.

in order to apply across situations and in order to achieve uniformity, in accordance with the rule of law. Litigants' hair colour is almost never a relevant consideration, and often motive or even ethical behaviour is not factored in either. Equity by contrast unabashedly employs context. Aristotle captured this aspect of equity when he likened equity to the leaden ruler used by the builders on the island of Lesbos; being soft, a leaden rule would mould around a stone and allow an exactly fitting stone to be chosen to fit next to it.³⁴ This process requires a kind of fine judgment.³⁵

Because it is a stand-by system, equity allows the law to be more formal. Instead of necessarily undermining the guidance function of the law, equity can allow the law to be simpler, more general and easier to follow – most of the time. The functional separation into law and equity also allows equity to target certain problems such as opportunism, without needless destabilisation. By using triggers based on proxies for problems like opportunism – fire alarms, if you will – equity is not hanging over every actor and every transaction all the time. While it is true that equity tends to look to custom and very basic consensus morality to keep its analysis within bounds, the use of triggers based on proxies means that we need not define opportunism with exactitude or build the limits of equity into the notion of morality or conscience.³⁶ Indeed, because it sounds in basic consensus morality, equity performs a kind of ‘acoustic separation’: with the same message, equity threatens the potential opportunist and reassures the garden-variety rule follower.³⁷ Again, by

³⁴ Bartlett and Collins (n 30) 112 [Aristotle, *Nicomachean Ethics* 1137b, lines 29–31].

³⁵ *ibid* 129–30 [1143a, lines 19–35].

³⁶ For examples of more direct definitions of equity, see D Klimchuk, ‘Equity and the Rule of Law’ in LM Austin and D Klimchuk (eds), *Private Law and the Rule of Law* (Oxford, Oxford University Press, 2014) 247–68; I Samet, ‘What Conscience Can Do for Equity’ (2012) 3 *Jurisprudence* 13.

³⁷ Y Feldman and HE Smith, ‘Behavioral Equity’ (2014) 170 *Journal of Institutional and Theoretical Economics* 137. Cf M Dan-Cohen, ‘Decision Rules and Conduct Rules: On Acoustic Separation in Criminal Law’ (1984) 97 *Harvard Law Review* 625. Emily Sherwin argues that equity may achieve acoustic separation in the sense of deceptively telling actors the law is rule-like and then actually applying fairness review: EL Sherwin, ‘Law and Equity in Contract Enforcement’ (1991) 50 *Maryland Law Review* 253, 300–14. See also EL Sherwin, ‘Equity and the Modern Mind’ in Goldberg, Smith and Turner (n 13). The acoustic separation I am arguing for does not involve two messages or any deception.

preventing evasion of the law, equity prevents it from unravelling: either by becoming too complex or falling into disrepute.³⁸

Finally, the idea that this stand-by system is second order also has points of contact with the historical tradition of equity. Aristotle's equitable function can be interpreted this way: the law need not make reference to equity, but equity does involve evaluation and intervention into the law. The early seventeenth century conflict between law and equity pitting Coke against Ellesmere solidified equity courts' ability to intervene from outside and to have the last word.³⁹ This jurisdictional structure dovetailed with a more substantive point: equity presupposed law and not vice versa. As Maitland pointed out, without equity 'in some respects our law would have been barbarous, unjust, absurd', but still 'the great elementary rights, the right to immunity from violence ... the rights of ownership [and so on] would have been enforced', but abolishing common law would have meant 'anarchy', because '[a]t every point equity presupposed the existence of common law'.⁴⁰ Maitland formulated this relationship in the famous aphorism that 'Equity without common law would have been a castle in the air, an impossibility'.⁴¹

The second-order aspect of equity is also reflected in how fusion happened. Generally speaking in both general legislation and more piecemeal forms of fusion the equitable element

³⁸ HE Smith, 'Property, Equity, and the Rule of Law' in Austin and Klimchuk (n 36) 224–46; M Harding, 'Equity and the Rule of Law' (2016) 132 *Law Quarterly Review* 278.

³⁹ See Baker (n 16) 125–26.

⁴⁰ FW Maitland, *Equity: A Course of Lectures* (AH Chaytor and WJ Whittaker eds, Cambridge, Cambridge University Press, 1936) 19. This view stands in contradiction to the views of those like Hohfeld who see equity as a complete and parallel system of law. See WN Hohfeld, 'The Relations Between Equity and Law' (1913) 11 *Michigan Law Review* 537, 557: 'Since, in every sovereign state, there must, in the last analysis, be but a single system of *genuine law*, since the various rules and principles of that system must be consistent with one another, and since, accordingly, all *genuine* jural relations must be consistent with one another, two conflicting rules, the one "legal" and the other "equitable", cannot be valid at the same moment of time: one must be valid and determinative to the exclusion of the other'.

⁴¹ *ibid.*

‘wins’ whenever the two conflict.⁴² This is true both substantively and procedurally, with the rise of notice pleading and free discovery, among many generalised equitable devices. Fusionists tend to see this as a one-time affair, but there should be a question of whether the ‘conflicting’ law is simply eliminated. If instead equity is a distinct function, one may be able to formulate the law as general enough to cover a large range of cases and then let equity override or suspend the law in some of them. If so, to the extent equity is dialed back the law automatically springs back to cover the situation in question. I will argue that sometimes this is exactly the right way to think about some equitable defences.

III. First Order Versus Second Order Defences

Defences have always constituted an important part of equity historically, and a functional account dovetails partially with the familiar picture and helps make partial sense of it. Equity both restrained litigants from enforcing their rights in an inequitable manner and imposed equitable limits on the special remedies that equity offered. Both can be thought of as defences, and in a fused system that is their natural place. Like the rest of equity, equitable defences often, but not always and not exclusively, served a safety-valve function as meta-law.

Regular parts of the law are first order, such as the operation of causes of action and defences like consent to battery or necessity to trespass. Defences like consent and necessity are rules or standards that operate as part of the ‘regular’ or first-order law: the defence requires no special comprehensive view of the law and relatively little discretion. The triggers for the defence are expressed in the same terms as the claims they defeat. Emily Sherwin notes that in contract, legal defences – ‘includ[ing] fraud, variations on fraud, certain types of mistake, incapacity,

⁴² For the United States, see Federal Rules of Civil Procedure (1938); New York Code of Procedure of 1848, NY Laws c 379 (Field Code); RE Kharas, ‘A Century of Law-Equity Merger in New York’ (1949) 1 *Syracuse Law Review* 186. For England, see Supreme Court of Judicature Act 1873; Supreme Court of Judicature Act 1875; Baker (n 16) 108–09; S Waddams, ‘Equity in English Contract Law: The Impact of the Judicature Acts (1873–75)’ (2012) 33 *Journal of Legal History* 185; PI McMahon, ‘The Fusion Fallacy Revisited: A Purposive Approach to the Fusion of Law and Equity Under the Judicature Acts, 1873–1875’ (unpublished manuscript, 30 Sept 2015).

duress, impossibility, and undue influence’ – tend to be ‘standardized’ and, compared to the ‘equitable fairness’ defence, ‘relatively narrow and concrete’.⁴³

A. General Considerations

In considering how equitable defences function as meta-law, we need first to show how conventional views of equity and equitable defences tend to reduce them to something first-order and thereby miss the equitable function that is our concern here.

In much of the law and economics literature, a great deal of what went under the heading of law versus equity is captured by rules versus standards. In Louis Kaplow’s model of the distinction, the difference is in the timing of decision-making.⁴⁴ Rules involve more ex ante specification than standards, and the latter leave more content to be filled in at the application stage. Thus, in the classic example, a flat speed limit is a rule, whereas a directive to drive reasonably under the conditions is a standard. (Actual laws tend to be a hybrid of the two.) The kinds of problems that call for special treatment by equity – opportunism, conflicting rights and multipolar problems – tend to require at least some use of standards. In Kaplow’s model, rules are more expensive to set up but cheaper to apply, whereas standards are the reverse: they leave most of the cost for the ex post time of application. The ‘equitable’ problems are ones in which it is difficult to come up with tailored ex ante rules, which pushes decision making towards standards. But both rules and standards can be either first or second order, and the two distinctions sometimes come apart in the face of the equitable problems. Thus, extreme opportunism, as in misfeasance by a fiduciary, calls for broad and simple ex ante rules and no filling in later through standards.⁴⁵ More usually, opportunism tends to defeat efforts to combat it on the same level. Ex ante rules are manipulable, and broad first-order standards are potentially very destabilising.

To see how going to a second level is different from replacing a rule by a standard, let’s return to the speed limit example. Software for adaptive traffic control and autonomous vehicles

⁴³ Sherwin, ‘Law and Equity in Contract Enforcement’ (n 37) 266–67 (footnotes omitted).

⁴⁴ L Kaplow, ‘Rules Versus Standards: An Economic Analysis’ (1992) 42 *Duke Law Journal* 557.

⁴⁵ Smith, ‘Why Fiduciary Law is Equitable’ (n 2).

that can build more fine-grained traffic rules into the system is being developed.⁴⁶ One potential problem is that human agents or software can anticipate the system and evade it, by performing dangerous self-serving manoeuvres that satisfy the constraints imposed by the system.⁴⁷ One response can be to plug the loopholes. Another would be to make the rules more standard-like but at the cost of greater uncertainty and slack in the system (more margin for error must be built in). Another alternative is to go one level up. This can be done by redesigning the system or building an extra layer of adaptation into it, by enabling the system to evolve in response to driver behaviour: the goal is for the system to make reference to primary rules and behaviour and change the primary rules in response to bad results stemming from opportunistic behaviour. This allows the system to make use of tightly interconnected rules and aim for closely meshing behaviour while at the same time addressing misuse of the system in a targeted fashion. It is the equitable function on wheels.

Going to a second-order level presents advantages but at the cost of significant disadvantages. Both the advantages and disadvantages form an answer to those who might argue that there is no difference between first-order and second-order law. It is a set of entwined advantages and disadvantages that is at stake. From a higher level, the higher-order components can control the variability at the primary level in a fashion that is comprehensive – it takes into account a wide range of information at the primary level – without thereby making that information available to all the first-order processes. By definition, first-order processes make use of whatever information they are formulated to take into account. They need not – but can be – formal, in terms of relative invariance to context.

⁴⁶ Thanks to Ted Sichelman for suggesting this example. See, eg, P Gora and P Wasilewski, ‘Adaptive System for Intelligent Traffic Management in Smart Cities’ in D Ślęzak, G Schaefer, ST Vuong and YS Kim (eds), *Lecture Notes in Computer Science, Book 8610 – Active Media Technology: 10th International Conference, AMT 2014, Proceedings* (Springer, 2014).

⁴⁷ For a humorous take on this problem, see ‘Engineers Unveil New Driverless Car Capable of Committing Hit-And-Run’ (*The Onion*, 2 April 2015) available at <http://www.theonion.com/article/engineers-unveil-new-driverless-car-capable-of-com-38358>, describing an ‘advanced Culpability-Evasion-System’.

In terms of equitable defences, such defences only operate when triggered, they make relatively free use of contextual information, and they (and that wide context) are not part of or referred to by the primary part of the law. Whether to formulate a defence as first order or second order involves a trade off familiar from systems theory. Setting up a second level gives rise to start-up costs and causes uncertainty relative to a system that does not make use of the contextual information it uses. There will also be uncertainty around the edges in whether the triggering condition for equity's second level has been met. And yet, the second level allows better tailoring – for example better containment of opportunism – than either a system that does make the attempt (no context, lots of opportunism) or a system that tries to accomplish everything on the same level. The latter can under some circumstances involve a greater degree of uncertainty than the uncertainty created by a second level of equity. If in principle it is all context all the time at the primary level, that can be more destabilising than triggered use of free context at an equitable meta level. There are all sorts of systems that combine different formulations of the two levels and their combination: how much context at level one, how much at level two, and what are the triggers. Which combination is best is an empirical question. For example, if opportunism is a very small problem, tolerating it would be less costly than using a lot of context at level one or level two. If the opportunism problem comes up rarely and can be dealt with by moderate amounts of context, formulating a standard at level one is a viable option. And so on. This is a problem quite familiar from systems theory, where all sorts of combinations of levels are employed in practice.

Returning to defences, we can now see how different first- and second-order defences are. First-order defences are conditioned on the same kinds of information and use the same kind of reasoning in their application as other parts of the law. Consider contributory and comparative negligence: each is conditioned on the same kinds of information as is the tort of negligence itself. Like the regular law, legal defences purport to offer guidance to actors' future behaviour. Thus, it makes sense to ask what the general effect on incentives of a given defence is, and the defence is no different in its terms than the claim it qualifies.

Consider first non-equitable, regular or, to use the historic (but not entirely accurate) term, 'legal' defences. Because they operate on the same level in this way, first-order defences are more likely to lead to controversy over whether they are defences at all: is the doctrine in question a defence or part of the main case? Is permission a defence to the tort of trespass to land or is lack

of permission part of the main case? In the United States, this is a surprisingly unclear proposition.⁴⁸

The defences that are functionally equitable are different. They are second order in the sense that they are in principle conditioned on the whole of the regular law and modify the result that would otherwise obtain. They are more discretionary and more context-dependent, and they make more reference to moral considerations. It is thus harder to confuse them with the main case. And, as we will see, this corresponds to one barrier to their fusion with the regular law: equitable defences differ in terms of how easily they can be recast at the primary ('legal') level. The easier that is, the easier it is both to fuse them with the law and to entertain the possibility that they are really not defences at all but part of the main case.

B. Functionally Equitable Defences

Some defences are second order, and reflect an equitable function. This is loosely, and only loosely, associated with historic equity jurisdiction. The separate courts did promote second-order defences, but second-order defences do not require separate courts. The traditionally equitable defences served a variety of second-order roles, which makes them differ in their degree of fusability with regular law.

i. Unconscionability

Illustrative of equitable defences that serve an equitable function is unconscionability. Unconscionability comes in many varieties, some of which are very narrow and others of which verge on a completely unconstrained fairness review. Equity courts treated unconscionability as a

⁴⁸ See, eg, *Environmental Processing Systems LC v FPL Farming Ltd*, 457 SW 3d 414, 425 (Tex 2015), holding that a plaintiff must plead and prove lack of authorisation; *Perkins v Blauth*, 127 P 50 (Cal 1912): '[W]e are not advised of any rule of pleading which requires a declaration from plaintiff that an unlawful trespass was committed without his acquiescence. There is no presumption that a plaintiff consents to an unwarranted invasion of his personal rights or rights of property'; *Moore v Walter Coke Inc*, No 2:11-cv-1391-SLB, US Dist ND (Ala 2012) 34, holding that a trespass plaintiff's burden regarding lack of authorisation was met by lack of facts pleaded suggesting consent.

prototypical instance of constructive fraud, involving advantage-taking,⁴⁹ especially of the vulnerable and, among modern accounts, those offered by Arthur Leff and Richard Epstein come closest to capturing the traditional approach, including (unconsciously) its second-order aspect.⁵⁰ The trigger for unconscionability is a combination of disproportionate hardship and a vulnerable party, with overtones of conscious behaviour or bad faith lowering the threshold. In terms of the specific proxies used under the heading of vulnerability, certain classes of people, such as the very young and very old, sailors on leave, and so on, triggered a presumption of opportunism.⁵¹ Once this threshold has been met, a more freewheeling fairness analysis with the presumption against the potential advantage-taker would ensue; if the transaction flunks, the contract is not enforceable, originally by specific performance and latterly also by damages. Epstein points to a similarity to the Statute of Frauds: some transactions contain a high enough degree of danger of being fraudulent that it makes sense not to enforce that class of transactions – even if some legitimate deals will be swept in the dragnet.⁵² The point is that banning a class of transactions can be better than doing nothing or further tailoring of the law itself. The difference with the Statute of Frauds is that equity is typically second order and, once triggered, it involves (even on Epstein’s somewhat narrow version) a more holistic and context-dependent analysis than is required under the rule-like Statute of Frauds.⁵³

⁴⁹ Story (n 19) § 221.

⁵⁰ AA Leff, ‘Unconscionability and the Code—The Emperor’s New Clause’ (1967) 115 *University of Pennsylvania Law Review* 485, 539; RA Epstein, ‘Unconscionability: A Critical Reappraisal’ (1975) 18 *Journal of Law and Economics* 293, 293–301. See also HE Smith, ‘The Equitable Dimension of Contract’ (2012) 45 *Suffolk University Law Review* 897, 902–07.

⁵¹ Leff (n 49) 532.

⁵² Epstein (n 49) 302.

⁵³ That is, the Statute of Frauds was rule-like until equity encrusted it with exceptions and modifications.

One might think – and many do think – that unconscionability is irrelevant for sophisticated parties.⁵⁴ One can grant that it is less needed in the case of sophisticated parties, but it is an open question whether it can ever be eliminated altogether. In such contexts, the possibility that a party will come up with an as yet unheard of method of committing near-fraud is not implausible. Perhaps the courts’ unwillingness to allow any parties to bargain away the duty of good faith is a recognition that courts need this ultimate trump card for such rare occasions – and that the benefits of legitimate instances of bargaining away do not outweigh the danger of opportunism that it presents elsewhere. Again, this is an empirical question, and invocation of parties’ sophistication will not resolve it. Nor does this represent a commitment of courts to root out opportunism wherever it might lurk.⁵⁵ As always, the question is whether the presence of the stand-by system is cost-effective or not, considering the possibility of error in both directions.

Consonant with its being a prototypical instance of constructive fraud, unconscionability should not be tied to equitable remedies. As we will see, there may be more need of second-order adjustment in the case of remedies like specific performance that lead to cliffs: beyond a certain point of primary activity the liability takes a sudden jump. And yet if one can commit constructive fraud – can engage in opportunism – through insistence on a damages remedy, the defence of unconscionability should be available there too. This is a case where a functional account of equity points to the desirability of fusion.

⁵⁴ See, eg, JS Kraus and RE Scott, ‘Contract Design and the Structure of Contractual Intent’ (2009) 84 *New York University Law Review* 1023; A Schwartz and RE Scott, ‘Contract Theory and the Limits of Contract Law’ (2003) 113 *Yale Law Journal* 541, 609–10.

⁵⁵ But see RE Scott, ‘Contract Design and the Shading Problem’ (2015) 99 *Marquette Law Review* 1.

ii. Estoppel

Estoppel is a prototypical equitable defence, in which a court will refuse to allow someone to create an expectation and then defeat it to the prejudice of the other party. Unlike unclean hands, estoppel can be used affirmatively to enjoin someone from enforcing a legal right.⁵⁶

Equitable estoppel is the effect of the voluntary conduct of a party whereby he is absolutely precluded both at law and in equity, from asserting rights which might perhaps have otherwise existed, either of property, of contract, or of remedy, as against another person, who has in good faith relied upon such conduct, and has been led thereby to change his position for the worse and who on his part acquires some corresponding right, either of property, of contract, or of remedy.

The situation of someone trying to benefit from this kind of inconsistent representation or behaviour is rife with the danger of opportunism. The triggers for estoppel are like those for unconscionability: a combination of extreme result, bad intent and vulnerability, the mixture of which varies by context (leading to an open question whether various forms of estoppel can be treated as one category).⁵⁷ It often requires a second-order approach: the net effect of the opportunism is hard to evaluate except in light of a wide range of interacting circumstances.

Like many equitable defences, estoppel allows for compromise between all or nothing in remedies. This is especially important for equitable remedies.⁵⁸ The injunction leads to large cliffs in liability: the difference between being ordered to do something under threat of contempt and not being ordered is not a matter of degree, despite the ability of courts to tailor or delay injunctions. Injunctions can lead to very lopsided results, leading to the defence of undue

⁵⁶ *Cunninghame v Cunninghame*, 772 A 2d 1188, 1201 (Md 2001), citing 3 J Pomeroy, *Equity Jurisprudence* 5th edn (Spencer Symonds, 1941) § 804. See also TL Anenson, 'The Triumph of Equity: Equitable Estoppel in Modern Litigation' (2008) 27 *Review of Litigation* 377.

⁵⁷ See E Cooke, *The Modern Law of Estoppel* (Oxford, Oxford University Press, 2000); MP Gergen, 'Towards Understanding Equitable Estoppel' in C Rickett and R Grantham (eds), *Structure and Justification in Private Law: Essays for Peter Birks* (Oxford, Hart Publishing, 2008) 319–37. See also B McFarlane and P Sales, 'Promises, Detriment and Liability: Lessons from Proprietary Estoppel' (2015) 1341 *Law Quarterly Review* 610.

⁵⁸ This is one of the main arguments Edward Yorio made for keeping equitable defences separate: see E Yorio, 'A Defense of Equitable Defenses' (1990) 51 *Ohio State Law Journal* 1201. See also HL McClintock, *Handbook of the Principles of Equity* 2nd edn (St Paul MN, West Publishing Co, 1948) § 23 (at 51); SL Bray, 'The System of Equitable Remedies' (2016) 63 *UCLA Law Review* 530.

hardship.⁵⁹ The on/off jump in liability and the potential lopsided results of equitable remedies like injunctions themselves carry a danger of opportunism. This is why the traditional rules of thumb for injunctions are sensitive to the potential for opportunism on both sides: the defendant may be in bad faith and the plaintiff may be an extortionist.⁶⁰

Estoppel is not limited to being a defence to equitable relief. It also relates to the use of equity to enjoin the enforcement of legal rights. As such, it is still second order, but this does not require it to be paired with an equitable remedy. Someone can opportunistically create reliance and disappoint it in such a way that the put-upon party would owe damages. For example, someone might lead another to believe that he will not enforce a right and then sue in damages for its violation.⁶¹

Much of what is special about estoppel is captured by regarding it as second order. Mark Gergen argues that mysterious statements that estoppel is a shield and not a sword may not have much content but they do reflect something: estoppel can be deep because it is narrow remedially.⁶² Within its narrow remedial ambit, estoppel involves broad evaluation, for example asking which party did more to create a risk.⁶³ These features of estoppel are consistent with estoppel being triggered as a second-order intervention and, once triggered, ranging broadly over the primary situation.

Although it is true that equitable remedies will often call for second-order defences, it is not the case that second-order defences are only of relevance to traditionally equitable remedies. When it comes to both traditionally equitable and traditionally legal remedies, we need to ask whether we can formulate triggers and engage in a worthwhile evaluation of the whole situation in order to solve problems like opportunism. Otherwise defences need not be functionally

⁵⁹ See sources cited at n 18.

⁶⁰ MP Gergen, JM Golden and HE Smith, 'The Supreme Court's Accidental Revolution? The Test for Permanent Injunctions' (2012) 112 *Columbia Law Review* 203.

⁶¹ *American Jurisprudence* 2nd edn (2016) vol 28, 'Estoppel and Waiver' § 27.

⁶² Gergen (n 56) 330–37.

⁶³ *ibid* 328–30.

equitable. When it comes to estoppel, defeating expectations to the detriment of another is a good proxy for the danger of opportunism, whether that trigger comes from the historic equity or law side of the old jurisdictional divide.

iii. Unclean Hands

Unclean hands is a defence related to a famous maxim of equity: ‘One who comes into equity must come with clean hands’ means that one cannot get an equitable remedy if one has acted immorally in the transaction in question. General bad behaviour or ill repute is not enough.⁶⁴ The unclean hands defence is of relatively recent vintage,⁶⁵ but like estoppel it allows a flexible and comprehensive view of the transaction in question. A useful counterpoint is the common law doctrine of *in pari delicto*, which operated more mechanically: it looked for equal guilt and invalidated the entire cause of action.⁶⁶

The hallmark of the unclean hands defence is flexibility. Unclean hands, like estoppel, can blunt the on/off quality in response to the overall situation. Unclean hands is neither a rule nor an all-things-considered first-order standard. It is triggered by bad behaviour in the transaction in question and freely makes use of context within that transaction. It is narrow and intense, like equity generally.

Perhaps the best argument for keeping unclean hands cabined in various ways is the difficulty of getting it right. Although there are not a great many cases in which someone manages

⁶⁴ *Scattaretico v Puglisi*, 799 NE 2d 1258, 1261–62 (Mass App Ct 2003): ‘A person is not to be deprived of civil justice merely because he has sinned in the past; his wrongdoing must have been related directly to the present situation to justify his being barred’. See also at 1262, fn 16: ‘Chief Baron Eyre who, according to Chafee ... first uttered the maxim, “A man must come into a Court of Equity with clean hands,” was well aware of the point: “it does not mean a general depravity; it must have an immediate and necessary relation to the equity sued for”’, citing *Dering v Earl of Winchelsea* (1787) 1 Cox Eq 318, 319; 29 ER 1184, 1185.

⁶⁵ Z Chafee Jr, ‘Coming into Equity with Clean Hands’ (1949) 47 *Michigan Law Review* 877, 880–84.

⁶⁶ TL Anenson, ‘Treating Equity Like Law: A Post-Merger Justification of Unclean Hands’ (2008) 45 *American Business Law Journal* 455, 489.

to get a legal remedy after being denied an equitable remedy for unclean hands,⁶⁷ one of the most famous equity cases involved just such a situation. In *Carmen v Fox Film Corp*,⁶⁸ actress Jewel Carmen repudiated a contract she had entered into just before the age of majority (invoking the legal defence of minority), and sought to sue the disappointed party for an injunction against interference with her new more lucrative contract. The court found the behaviour in question to be unethical even if legal, and denied the injunction. Carmen later sued and obtained substantial damages. The case is set against the backdrop of the collusion among movie companies in the studio system of early and mid twentieth century Hollywood. Perhaps, courts are not well equipped to deal with a problem on that scale, and so contract enforcement should not be conditioned on it. It is even possible that preventing Carmen from getting an injunction but allowing damages is the least bad response to an opportunist acting in a sea of opportunism. The right question to ask is what the best rules of thumb are and at what level of generality. The current state of the rules of thumb is to allow unethical behaviour more easily to block access to the most troubling remedies, and to allow for a more continuous type of modulation for damages.

iv. Laches

In laches, a defendant can get a claim dismissed because of the plaintiff's unreasonable and prejudicial delay in filing suit, even if the statute of limitations has not yet run. Traditionally, the defence of laches only precluded equitable relief – such as injunctions, specific performance and accounting – but would not affect legal remedies like damages.⁶⁹ Recently, a few courts have extended laches to legal claims.⁷⁰ From a doctrinal point of view, there is good reason to apply laches to equitable remedies, even under recent federal legislation containing statutes of limitation. The extension to legal claims is more difficult, and I will consider it from a purely normative point

⁶⁷ Sherwin, 'Law and Equity in Contract Enforcement' (n 37) 259, fn 26 (citing cases).

⁶⁸ 269 F 928 (2d Cir 1920).

⁶⁹ *Landreth v First National Bank of Cleburne County*, 45 F 3d 267, 271 (8th Cir 1995); *Erwin v City of Palmyra*, 119 SW 3d 582, 586–87 (Mo App ED 2003).

⁷⁰ *Harris v Beynon*, 570 F Supp 690, 692, fn 3 (ND III 1983); *Sutton v Davis*, 916 SW 2d 937, 940–41 (Tenn Ct App 1995). See also E Fetter, 'Laches at Law in Tennessee' (1997) 28 *University of Memphis Law Review* 211.

of view. (This is the point of view that is available to English courts and most American state courts under state law, and would be available under US federal law but for a narrow view of the nature of equity at that level.⁷¹)

Laches clearly takes more context into account than does a statute of limitations: it is a standard and not a rule. Like estoppel, it can also be used to police for opportunism, and some of its applications can involve hard-to-foresee and unfair advantage-taking. The latter make laches look more second order.

Should laches apply to legal remedies? Laches originally was equity's response to a lack of statutes of limitations.⁷² Modern statutes of limitations are often argued to reflect a legislature's considered judgment of a reasonable time for bringing suit – full stop. And yet, it is not impossible for someone to engage in opportunism in lulling another into reliance on the prospect of not being sued in damages.

As with other equitable defences, it is not hard to see why they not only arose in equity but are most compelling when it comes to equitable remedies. As Sam Bray points out, the equitable remedies are often burdensome on the enjoined party, require management of the parties and are particularly costly, whereas damages often (but not always) involve a sliding scale.⁷³ Again, the on/off quality and large cliffs argue for stepping back in a second-order way. Indeed, the whole theme of 'management' of the parties and of the litigation that Bray identifies in equity is naturally performed at a meta level. Bray argues further that the equitable defences like laches are costly in terms of uncertainty and especially compelling for equitable remedies such that drawing the line

⁷¹ The US Supreme Court looks to what courts of equity did in 1789: *Grupo Mexicano de Desarrollo SA v Alliance Bond Fund Inc*, 527 US 308 (1999), and see SL Bray, 'The Supreme Court and the New Equity' (2015) 68 *Vanderbilt Law Review* 997.

⁷² *Petrella v Metro-Goldwyn-Mayer Inc*, 134 S Ct 1662, 1673 (2014), citing 1 D Dobbs, *Law of Remedies* 2nd edn (St Paul MN, West Publishing Co, 1993) § 2.4(4) (at 104): 'laches ... may have originated in equity because no statute of limitations applied ... suggest[ing] that laches should be limited to cases in which no statute of limitations applies'.

⁷³ SL Bray, 'A Little Bit of Laches Goes a Long Way: Notes on *Petrella v. Metro-Goldwyn-Mayer Inc.*' (2014) 67 *Vanderbilt Law Review En Banc* 1.

at the law-equity divide makes sense. In the recent case of *Petrella v Metro-Goldwyn-Mayer Inc.*,⁷⁴ the US Supreme Court took this middle route – reserving laches for equitable remedies – and not extending the defence to legal remedies as urged by leading pro-fusion scholars.⁷⁵

Moreover, as many have noted, for the more extreme cases of opportunistic delay of suit within the statute of limitations, the defence of estoppel is uncontroversially available.⁷⁶ Because estoppel directs the court to a more explicitly second-order and constrained equitable analysis and a generalised laches defence would invite a great deal of uncertainty, it is reasonable to cabin laches to equitable remedies and rely on estoppel for the worst abuses. This is not because damages and other legal remedies never present problems that could benefit from ‘meta’ treatment or that some applications of laches could not be performed at the primary level. Rather, it would seem that the current system gets much of the envisioned benefit without courting the all-too-familiar danger of equity filling the entire space. Given that second-order power is not easy to cabin, we should ask for more than the tidiness of complete fusion or a ‘why not’ style presumption before extending laches further. Unlike in the pro-fusion commentary, the burden at this point should be on those advocating extension.

v. Fraud and Constructive Fraud

Many defences appear to be a little ambiguous as to whether they are functionally legal or equitable, and this is especially true after the fusion of law and equity. A microcosm of this picture comes from fraud in contract.⁷⁷

⁷⁴ *Petrella* (n 71). *Petrella* involved a copyright claim. The Court has recently decided that laches does not apply to claims for damages from patent infringement during the statutory period under 35 USC § 286: *SCA Hygiene Products Aktiebolag v First Quality Baby Products*, 137 S Ct 954 (2017).

⁷⁵ See ‘Brief of D Laycock, MP Gergen, and D Rendleman as Amici Curiae in Support of Neither Side’ (2013) WL 6213269 (US, Appellate Brief) and ‘Brief of TL Anenson as Amicus Curiae in Support of Respondents: *Petrella v Metro-Goldwyn-Mayer*’ No 12-1315.

⁷⁶ Bray (n 72) 8.

⁷⁷ Contractual mistake presents a similar picture, with a broader and more complex approach to mistake in equity than at law: McClintock (n 57) § 88 (at 238). This has given way to an extension of mistake to claims for legal remedies:

Although the relation between legal and equitable fraud is a large subject – too large a subject for thorough treatment here – aspects of this relationship are very telling from the point of view of the equitable function. As already discussed, equity’s notion of constructive fraud was not just wider but more open-ended and the devices for dealing with it are structured as a second-order safety valve.

From the point of view of fusion, the types of fraud present a complex picture. At first blush, it looks as if law and equity are concerned with the same thing and that this area is ripe for reconciliation and consolidation. And yet fraud is not all of one piece. Some of it is easy to specify – lying to get someone to give you their property and the like – but some of it is highly uncertain. The main idea of constructive fraud is the difficulty of proving and especially anticipating new ways of abusing and evading the law.⁷⁸ This suggests that some residuum of second-order intervention will be needed. And yet relying only and always on second-order intervention would be needlessly uncertain. Again, a division of labour even within the domain of fraud carries its advantages.

Definitions of fraud have shifted over time as well. This too is no argument against an equitable function. Quite the contrary. As the equitable function is applied to categories of fraud, they become known and amenable to first order (‘legal’) treatment. Thus, yesterday’s purely equitable fraud sometimes becomes today’s legal fraud, notably in the law’s late recognition of fraud in the inducement, first dealt with in equity.⁷⁹ This pipeline of fraud from equity to law is an instance of the temporal dimension of the equitable function. It is to this dynamic aspect that we now must turn.

Restatement (Second) of Contracts (1981) ch 6, Introductory Note, 152–55. In English law, mistake may in cases of great hardship lead to a refusal of an injunction but allowance of damages. There are, however, no special equitable grounds to rescind contracts on grounds of mistake, a contract being either valid or invalid. See McGhee (n 19) para 15-002.

⁷⁸ See above n 22 and accompanying text.

⁷⁹ DJ Ibbetson, *A Historical Introduction to the Law of Obligations* (Oxford, Oxford University Press, 199) 208–09.

C. Patterns of Equitable Defences

Looking at the defences through the lens of an equitable function allows us to discern a number of patterns in defences.

First, we should expect that equitable defences will tend to respond to opportunism, polycentric problems and other complex and hard-to-foresee situations. From a functional point of view, this means that we should see a tendency for these problems to be solved with second-order interventions set off by triggers, especially those relating to disproportionate hardship and bad faith. There is nothing historical or jurisdictional about this, nor is it peculiar to the ‘common law’ world. And indeed civilian doctrines like abuse of right, abuse of law and good faith bear some resemblance to the equitable interventions and defences, from a functional point of view.⁸⁰

A correlation of such defences with historic and jurisdictional equity is not unexpected though. Because equity courts did intervene from outside and won the battle with the law, and because they focused on problems that such a tool was suited for, it is no accident that second-order safety-valve-style intervention was the order of the day for equity courts.

More specifically, where a defence is both legal and equitable, the equitable version can be expected to be more second order and to respond to those kinds of problem. This is simply a corollary of the general distinction between defences. This apparent duplication need not be duplication at all. Instead, the legal and equitable versions of the defences may be serving different functions in different ways. Fraud and constructive fraud cannot be collapsed into each other.

We also should expect some – but only some – correlation of traditional equitable defences and equitable remedies. As many have noticed, equitable remedies can lead to extreme results and equitable defences allow for compromise.⁸¹ To this we can add that the method for doing so is targeted, in a second-order fashion.

⁸⁰ See, eg, A di Robilant, ‘Abuse of Rights: The Continental Drug and the Common Law’ (2010) 61 *Hastings Law Journal* 687; LM Katz, ‘Spite and Extortion: A Jurisdictional Principle of Abuse of Property Right’ (2013) 122 *Yale Law Journal* 1444; JM Perillo, ‘Abuse of Rights: A Pervasive Legal Concept’ (1995) 27 *Pacific Law Journal* 37.

⁸¹ See above n 57.

When it comes to fusion, equitable defences are special. The ones that do indeed serve a second-order role should be hardest to fuse completely, and this does seem to be the case. Alternatively, when they are fused they will lead to the most complication or uncertainty. The reluctance of courts to fuse unclean hands and laches may receive some explanation here. The attempt to capture estoppel with proliferations of multifactor tests is a possible example of the latter.⁸² In other common law countries, fusion has led to a combination of standards and complex rules – what we would expect from an attempt to replace equity with something more first order.⁸³

IV. The Dynamics of Equitable Defences

Legal and equitable defences do not stay that way. Equitable defences, like other aspects of equity, have often been adopted by the common law courts and have become regular law.⁸⁴ Fusionists sometimes take this as further evidence of the arbitrariness of the law-equity distinction: any such distinction in the area of defences must be a happenstance, path dependence or the like.

It is actually this fusionist conclusion that is too static. This development of equitable into legal defences reflects the operation of the equitable function over time. First note that the traffic was one way: equitable defences and other equitable interventions became law but not vice versa. This was widely recognised by courts, especially as equitable treatment of mortgages became a matter of law.⁸⁵

⁸² Smith, ‘Fusing the Equitable Function’ (n 2). On multifactor tests for estoppel, see Gergen (n 56) 325.

⁸³ The multifactor test seems to be particularly prevalent in the United States. In jurisdictions without a tradition of Realism, there is a tendency to combine standards with finely tuned complex doctrine. See J Getzler, ‘Patterns of Fusion’ in P Birks (ed), *The Classification of Obligations* (Oxford, Clarendon Press, 1997) 192–93. This too can be seen as an attempt to do on one level that which, prior to fusion, was done on two.

⁸⁴ See, eg, *Spect v Spect*, 26 P 203, 205 (Cal 1891) (Harrison J), quoting Lord Redesdale; Anenson (n 65) at 463–64; RS Stevens, ‘A Plea for the Extension of Equitable Principles and Remedies’ (1956) 41 *Cornell Law Quarterly* 351, 352, 354.

⁸⁵ *Spect* (n 83) 205 (Harrison J), quoting Lord Redesdale.

The distinction between strict law and equity is never in any country a permanent distinction. Law and equity are in continual progression, and the former is constantly gaining ground upon the latter. A great part of what is now strict law was formerly considered as equity, and the equitable decisions of this age will unavoidably be ranked under the strict law of the next.

The statutes effecting fusion also gave primacy to equity whenever there was a conflict.⁸⁶ Commentators too have noticed this progression, especially in the area of defences.⁸⁷

Looking at these developments through the lens of equity as a second-order function helps explain and justify some of these dynamic patterns. First of all, the one-way traffic is what we would expect: equity is a power ranging over the law so that its output can be formulated in a first-order fashion. If for example, interventions to prevent murders by heirs – using the principle that one will not be allowed to profit from his own wrong – are recognised as a class of problems, a rule (the slayer rule) can replace the equitable intervention.⁸⁸ Likewise, when varieties of fraud (notably fraud in the inducement) were recognised as being amenable to first-order treatment they could become ‘legal’. It is hard to see how or why on the two-tier model that legal defences would become equitable in the same fashion. Further, Rose’s model of oscillation is not universal; sometimes the mix of law and equity (in a functional sense) can be stable, where aspects of equity become regularised and a local equilibrium is reached, in a process I call ‘sedimentation’.⁸⁹ Arguably this has happened with the role of notice in real property recording acts and more generally with good faith purchaser rules.

⁸⁶ See the sources cited above at n 41.

⁸⁷ JB Ames, ‘Specialty Contracts and Equitable Defenses’ (1896) 9 *Harvard Law Review* 49; Anenson (n 65) at 463–66. Even in Carol Rose’s model of oscillation between ‘crystals’ and ‘mud’, bright-line law (‘crystal’) becomes less determinate and more standards-based (‘mud’) through judicial interpretation and modification; courts’ mud is replaced with crystal by legislatures and parties, not by courts. CM Rose, ‘Crystals and Mud in Property Law’ (1988) 40 *Stanford Law Review* 577, 588.

⁸⁸ The chestnut in the United States is *Riggs v Palmer*, 22 NE 188 (NY 1889), which I have argued is an example of hyperfusion. See Smith, , ‘Fusing the Equitable Function in Private Law’ (n 2). The Wills Act was probably passed against a background of equity and earlier English ecclesiastical law which would have reached the same result.

⁸⁹ HE Smith, ‘Rose’s Human Nature of Property’ (2011) 19 *William & Mary Bill of Rights Journal* 1047, 1047–55.

To the extent it reflects an equitable function, equity is thus a moving frontier. The function is needed for problems of great uncertainty and variability. The same ‘problem’ need not be static in terms of its uncertainty and variability: especially if it becomes more certain, the equitable function need not be used on an ongoing basis, or it can be reserved for the residuum of the problem that retains its uncertain and variable character.

Far from providing an argument for the arbitrariness and hence undesirability of distinct equity, this dynamic picture provides powerful evidence for recognising an equitable function, both descriptively and normatively. The patterns themselves are what one would expect from the use of the equitable function over time and the building up of experience with applications. From a normative point of view, it makes sense to treat categories of problems in a less complicated way at the primary level if they are amenable to primary-level treatment. This is what we mean by the presumption for the first level (roughly, law) and the exceptional character of the equitable second level.

V. Conclusion

Fusionism often calls for refashioning defences according to their function. In this chapter I have suggested that recognising a second-order safety-valve function makes some sense of the traditional category of equitable defences. Because the equity courts were structured as separate and they wound up in a position to override the law, it was natural for them to focus on this equitable function. This is not all they did, however, and the law itself contains within it instances of second-order functions as well. It would be a mistake to draw the conclusion that freezing the jurisdictional line is any way justified. And yet, the rough allocation of defences to one side or the other, as with lack of capacity and statutes of limitation on one side and unconscionability and laches on the other, serves some purpose and should not be lightly tossed aside. Further, the idea that some defences straddle the line or are in the process of being adapted in a more formal way – as with fraud over a long period of time – should be welcomed and accepted. A second-order function can in principle be kept separate enough without separate courts. And yet, in a legal culture where the only boundaries to the judicial function are the brightest of bright lines with everything else up for grabs is not one that is very hospitable to equity. And in the United States,

the process of replacing equity, including the equitable defences, with first-order standards, including multifactor tests, is well under way. The functional account of equity offered in this chapter implies that we should call these developments into question rather than celebrating them unthinkingly as the pure dawn of enlightenment. The lessons of systems theory suggest that we will never get away from some version of equity, and certainly not the problems it is suited to address.