Exploring justice in extreme cases: Criminal law theory and international criminal law

Robinson, D.E.

04

Fundamentals Without Foundations

OVERVIEW

In Chapter 3, I discussed why we may need to reconsider familiar formulations of fundamental principles when we apply them in new contexts. For example, what does the legality principle require in a system without a legislature? What does the culpability principle require in contexts of collective violence? In this chapter, I ask how we might even embark on such evaluations. How would we go about formulating and evaluating the principles themselves?

A typical and commendable scholarly reflex would be that we must 'ground' our analysis in a secure foundation. In other words, we should be able to show that each proposition is justified by deeper premises, and those premises if challenged should in turn be demonstrably justified, until we reach a bedrock that is certain and self-evident or agreed by all. In this way, we would know we have reached the 'correct' deductions.

In this chapter, I show the infeasibility of rooting fundamental principles in secure moral foundations. We do not have an uncontroversially 'correct' foundational moral theory, and furthermore the comprehensive moral theories tend to lack the precision to dictate answers to most granular problems. Happily, this absence of bedrock does not mean that we must abandon thoughtful, rigorous discussion of fundamental principles.

I suggest a non-foundational approach, using a coherentist method: we do the best we can do with the available clues and arguments. The clues include patterns of practice, normative arguments, and casuistically-tested considered judgments. We can work with 'mid-level principles' (principles intermediate between practice and foundational theories) to carry out fruitful analytical and normative work.

The coherentist approach accepts that our principles are human constructs, that our starting points are contingent, and that we have no guarantees of 'correctness'. Nonetheless, it is important to *try* to determine whether institutions are just, using the best available methods that we have. Discussion of fundamental principles is not a matter of ethical computations; it is a *conversation.* It is a *human* conversation, a *fallible* conversation, and nonetheless an *important* conversation. I also argue that coherentism offers the best explanation of the method of most criminal law theory: in other words, it is the best theory of criminal law theory.

4.1. TERMS: FUNDAMENTALS AND FOUNDATIONS

I should explain some terms. I use the term 'fundamental principles' in the same way it is used in criminal law scholarship and jurisprudence: principles such as legality or culpability that are found to be fundamental within the legal system. The principles are 'fundamental' in comparison with other rules and doctrines in the system. By 'foundations' I mean ultimate bedrock justifications for beliefs; in ethical discourse the term is also used to refer to general comprehensive moral theories. My point is that we can make meaningful progress in discussing and refining fundamental principles of a criminal justice system without resolving ultimate moral questions, without necessarily subscribing to one of the main comprehensive theories, and without having to decide which comprehensive theory is the 'right' one.

I also use the term 'mid-level principle', but in doing so I am not drawing a hierarchy between 'fundamental' and 'mid-level'. I am simply adopting terms used in two bodies of literature. In ethics literature, the term 'mid-level principles' refers to principles that are arguably immanent within a body of practice. Mid-level principles are analytically useful, because they help explain and systematize the practice, and also normatively convincing. They are 'mid-level' because they mediate between legal practice and the foundational moral theories; they are more general than the former and more concrete than the latter. My argument is that fundamental principles of criminal justice, and our specific formulations of those principles, can be fruitfully analyzed as 'mid-level principles' in this broader sense.

4.2. WHERE CAN WE FIND FUNDAMENTAL PRINCIPLES?

As I discussed in Chapter 1, there are two distinct reasons why it is valuable to formulate the constraining principles as best we can. First, if we neglect or *understate* a fundamental principle of justice, we breach a commitment of fair treatment owed to the individual: we are treating the person unjustly. Second, and conversely, if we *overstate* a fundamental principle, we are being unnecessarily conservative; we are sacrificing social desiderata when no deontic constraint requires us to do so. It is 'bad policy', because we are failing to fulfill the societal aims of the system for no reason. Thus, we have both deontological and consequentialist reasons to develop plausible accounts of fundamental principles: it can help avoid unjust treatment, and it also helps develop better policy.

But where do look to find those principles? The literature commonly refers to two principles – culpability and legality – but how do we know to accept those two principles? How would we determine if there might be others? Where do we turn to see how to formulate their specific requirements? By 'formulations', I mean the articulations of specific implications. For example, does the legality principle require written legislation or can other notice suffice?[1] Does the culpability principle require some causal contribution to a crime, and if so, how much contribution is enough?[2]

At present, when ICL literature invokes or articulates a principle, it draws on any of three sources of reference: (1) formulations in *ICL authorities*; (2) induction from *national legal systems*; and (3) deduction from *philosophical argument.* Each of these three sources of reference is routinely invoked in ICL scholarship and jurisprudence. The way that all three are freely invoked may at first seem haphazard, but I will suggest below that the recourse to these reference sources is justified and appropriate, and that coherentism is actually the best explanation for how these reference sources are employed.[3]

First, however, in order to show that there are no simple and certain sources or methodologies available to us, I will inspect each source in isolation to demonstrate that each has strengths and weaknesses. We are going to see two recurring problems. One problem is the tradeoff between positivity and normativity. By 'positivity', I mean

[1] Chapter 5.
[2] Chapter 6.
[3] See below § 4.4.

recognized legal applicability and ascertainability, and by 'normativity', I mean the degree of convincingness that we 'ought' to recognize the principle. The second problem is the lack of a reliable foundation even for a purely normative conversation.

For greater certainty, let me specify that the methodology I propose below does not purport to *escape* all of these problems. Rather, I am simply demonstrating need for a methodology that openly acknowledges and responds to these problems. I will advocate a method that embraces non-certainty and non-simplicity, that strives to identify and test the weaknesses inherent in each source of reference, and that gives us tools for helpful deliberation despite these challenges.

4.2.1. First Source of Reference: Internal Formulations

The first source of reference is to use the articulations of principles as recognized in ICL jurisprudence itself (i.e. in its legal instruments and judicial pronouncements). It is an internal and doctrinal approach. For example, if we were debating the culpability principle, we might turn to the articulation of the principle in Tribunal jurisprudence, i.e. 'a person can only be held responsible for a crime if he contributed to it or had an effect on it'.[4] Or, if we were debating the requirements of non-retroactivity and the legality principle in an ICC case, we might invoke the terms enshrined in the ICC Statute.[5]

The strength of this approach is its 'positivity'. The principles are clearly legally applicable in ICL, because ICL itself says so. They are also relatively concrete, because we use the articulations provided in the authoritative pronouncements of ICL sources.

The weakness of this source is its limited normativity. The approach does not help at all with the question of whether ICL has adopted flawed or problematic understandings, or whether it ought to recognize other principles. In other words, we cannot use ICL understandings as a yardstick to *critically evaluate* ICL understandings. A purely internal approach also does not help us in liminal cases, where we need to further

[4] *Prosecutor v Tadić,* Judgement, ICTY A.Ch, IT-94-1-A, 15 July 1999 ('*Tadić* Appeal Judgement') at para 186; *Prosecutor v Kayishema,* Judgement, ICTR T.Ch, ICTR-95-1, 21 May 1999, para 199.
[5] A person can only be held responsible for a crime that was, at the time of its commission, a crime within the jurisdiction of the ICC: Article 21-23, ICC Statute.

specify how a particular principle should be formulated. We need some external framework for these kinds of evaluations.[6]

Furthermore, a purely internal account cannot tell us how to resolve a conflict between a doctrine and a principle. From the standpoint of formal non-contradiction, a conflict between an ICL doctrine and an ICL principle can be resolved by reforming the doctrine, or by re-formulating or even rejecting the principle.[7] Internal non-contradiction does not tell us whether to reform the doctrine or to reconsider the current understanding of the principle.

The internal approach is nonetheless analytically valuable, because it can reveal *internal contradictions* between ICL doctrines and principles. Internal non-contradiction is an important value in its own right, and thus internal contradictions should be detected and corrected.[8] Nonetheless, we need some external benchmark in order to *specify* the recognized formulations, to *adapt* them, or to critically *evaluate* them.

---

[6] To be more precise, the framework must be at least *partly* external, in order to be able to question critically the internally adopted formulations. The account I propose below draws on both internal and external inputs. Furthermore, while simple *consistency* does not tell us which way to redress a conflict, I argue below that broader *coherence* gives significantly more guidance.

[7] As an illustration, consider for example the culpability principle and the command responsibility doctrine. Early ICTY jurisprudence went from the following premises to the following conclusion:

1. *ICL respects the culpability principle, which requires that the accused must contribute to or have an effect on a crime to share in liability for it.*
2. *However, under the command responsibility doctrine, the accused need not contribute to or have an effect on a crime to share in liability for it.*
3. *Therefore, the culpability principle's requirement of causal contribution apparently does not apply to command responsibility.*

Whereas I would have reached a different conclusion:

3. *Therefore, we should re-examine the command responsibility doctrine to bring it into conformity with the culpability principle.*

There are of course many subtle details to the command responsibility debate, and you might disagree with how I characterize the ICTY analysis. I will unravel all of that with great care in Chapter 6 (including the 'separate offence' characterization that later emerged). The point I am making here is simply that, on a pure *internal consistency* account, the first solution is not 'wrong'. Cases like *Čelebići* assumed rather insouciantly that there must be an exception within the fundamental principle, but in doing so they removed the apparent conflict, at least from the formal logical perspective of internal consistency.

You may object that, even on an internal consistency account, there is still a formal problem with the Tribunal's chain of reasoning. Namely, it misunderstood the proper 'hierarchy' between fundamental principles and doctrines when it took the doctrine as the fixed point and assumed an exception in the principle. That could be a correct critique of the *reasoning* in this instance. But the *outcome* of re-interpreting the principle is not necessarily always wrong, if it is done after careful deontic analysis.

[8] Of course, any dynamic living legal system is a field of contestation, absorbing new values over time, and thus, contradictions will arise incidentally as components change. Nonetheless, contradiction cannot be a desideratum of any legal system, and coherence must be a systemic goal.

4.2.2. Second Source of Reference: Induction from National Systems

The second source is of reference to derive principles and their formulations by induction from the national legal systems of the world. ICL jurisprudence often canvasses national systems for guidance; for example, in the *Erdemović* case at the ICTY, the judges surveyed national systems to see if there was a common approach as to when duress is an excusing condition.[9] In the *Lubanga* case, chambers adopted the 'control theory' as a basis to distinguish principals and accessories, noting inter alia that the approach is applied in numerous legal systems.[10] Scholars and jurists frequently employ induction from national systems when articulating fundamental principles.[11]

This approach has *intermediate* levels of positivity and normativity. As for positivity, the 'general principles of law derived… from… legal systems of the world' have recognized legal applicability: they are a well-accepted subsidiary interpretive source of ICL.[12] This technique, adapted from general international law,[13] involves a survey of national systems to identify commonalities that can then guide the international system. General principles also offer some level of concreteness, since there is a broadly agreed methodology drawing on objectively ascertainable data.

---

[9] *Prosecutor v Erdemović,* Judgement, ICTY A.Ch, IT-96-22-A, 7 October 1997 ('*Erdemović* Appeals Judgement') and see more detailed discussion below, §5.2.2. In that case, there was too much discrepancy between national approaches to extract a general principle on the issue in dispute (whether duress was available for murder).

[10] See eg. *Prosecutor v Thomas Lubanga Dyilo*, Decision on the Confirmation of Charges, ICC PTC, ICC-01/04-01/06, 29 January 2007, para 330.

[11] Examples of drawing on national systems for guidance are innumerable; as illustrations see S Dana, 'Beyond Retroactivity to Realizing Justice: A Theory on the Principle of Legality in International Criminal Law Sentencing' (2009) 99 Journal of Criminal Law and Criminology 857 esp at 879-881, canvassing national approaches to *nulla poena sine lege*; K Gallant, *The Principle of Legality in International and Comparative Criminal Law* (CUP, 2009); K Ambos, *Treatise on International Criminal Law, Volume I: Foundations and General Part* (OUP, 2012) at 88 (legality) and 94 (culpability).

[12] See eg. ICC Statute, Art. 21(1)(c); '*Erdemović* Appeals Judgement', Opinion of Judges McDonald and Vohra, paras. 56-72. The ICC Appeals Chamber has held affirmed the value of drawing inspiration from national legal systems: '[T]he Appeals Chamber considers it appropriate to seek guidance from approaches developed in other jurisdictions in order to reach a coherent and persuasive interpretation of the Court's legal texts. This Court is not administrating justice in a vacuum, but, in applying the law, needs to be aware of and can relate to concepts and ideas found in domestic jurisdictions.' *Prosecutor v Thomas Lubanga Dyilo*, Judgment on the Appeal of Mr Thomas Lubanga Dyilo against his Conviction, ICC ACh, ICC-01/04-01/06-3121-Red, 1 December 2014. In that particular case, the method was not a general survey of systems, but rather a reference to German legal thinking, on the grounds that it offered a convincing normative theory that fit well with the ICC Statute and with the nature of the crimes before the Court, as a basis to distinguish principals and accessories. This approach matches the coherentist method discussed below in this chapter.

[13] ICJ Statute, Article 38(1)(c).

However, the positivity is only 'intermediate', because there are also significant limits. The immense difficulty of collecting the necessary data reduces the accessibility of this source. More problematically, the process requires that one *generalize* from many differing approaches of national systems, which reduces the specificity of the general principles and thus their concreteness in resolving specific issues. This is particularly a problem if national systems diverge on the precise question one is trying to resolve.[14]

Induction from national systems also has an intermediate level of *normativity*. On the one hand, there are good reasons to accord some normative weight to principles derived from national systems. After all, principles recognized across regions, cultures and traditions, and worked out based on decades or centuries of experience, offer an excellent guide to widely-shared intuitions of justice. They are a valuable reference point in informing our understandings of the proper constraints of the criminal sanction.[15]

On the other hand, there are also limits to that normative weight. One problem is that some national traditions will get 'double-counted' insofar as they exported their legal systems through colonization.[16] Thus, we cannot automatically assume that national principles reflect local intuitions of justice; we must be ready to examine biases and impositions of power that may have led to the predominant formulations.[17] Another problem is that, as discussed in Chapter 3, ICL operates in contexts that are often profoundly different from the 'normal' societal context in which the familiar formulations of principles evolved. So, even if every system in the world concurred in a particular formulation of a principle, that would not necessarily be a conclusive case for its absorption into ICL, because there could be morally salient differences. For example, even if every legal system in the world said that all criminal law must be written law,[18]

---

[14] In addition to the limits on positivity noted here, general principles are only a subsidiarity interpretive source, ranking below an institution's basic instrument as well as any relevant treaty and custom. See e.g. Art 21(1)(c) ICC Statute, and see *Erdemović* Appeals Judgement, above.

[15] G Fletcher, *The Grammar of Criminal Law: American, Comparative and International,* Vol 1 (OUP, 2007) at 66-67 and 94 (comparative study can inform our philosophical inquiry; attempt to formulate principles that cut across legal systems; avoid parochialism); K Ambos, 'Toward a Universal System of Crime: Comments on George Fletcher's *Grammar of Criminal Law*', 28 *Cardozo L Rev* (2006-7) 2647 at 2647, 2649 and 2672; P H Robinson and R Kurzban, 'Concordance and Conflict in Intuitions of Justice', (2006-7) 91 *Minnesota Law Review* 1829.

[16] J Stewart and A Kiyani, 'The Ahistoricism of Legal Pluralism in International Criminal Law' 65 *American Journal of Comparative Law* (2017) 393.

[17] Stewart and Kiyani, 'Ahistoricism', ibid.

[18] They do not. Consider for example the UK, which still allows common law offences, as well as the many local regimes of customary law in the world. See e.g. D D Ntanda Nsereko, *Criminal Law in Botswana* (Kluwer, 2011) eg at 46; D Isser, ed, *Customary Justice and the Rule of Law in War-Torn Societies* (USIP Press, 2011).

that would not necessarily support a maxim that all criminal law systems in all circumstances must be based on written law. We might conclude instead that *lex scripta* is merely a manifestation of a deeper underlying principle (perhaps concerning notice or ascertainability), and the specific requirement of written legislation applies only under certain societal conditions. ICL, at least in its early phases, provides an interesting context to explore the possible unstated preconditions of the *lex scripta* requirement.[19]

Thus, the second source, induction from national systems, offers intermediate positivity and normativity. National formulations provide some guidance to widely-shared understandings of justice, worked out over time in diverse settings. However, we must be alert to possible biases in existing practice and formulations of principles, as well as possible inapplicability outside the familiar societal contexts in which those formulations were developed.[20]

4.2.3. Third Source of Reference: Deduction from Moral Philosophy

The third source of reference is to deduce conclusions from normative argumentation; for example, by appealing to basic moral commitments as to how persons should be treated. This is also a frequently-employed method; for example, when one invokes philosophical thinkers or school of thought,[21] or when scholars or jurists engage directly with moral questions of justice for the individual.[22]

This approach has the highest level of normativity, since it is purely a discussion about what we *ought* to do. The obvious problem with the third approach is, of course,

[19] For further discussion, see Chapter 5.

[20] For discussion, see Chapter 3.

[21] Examples abound, but one illustration among many would be George Fletcher drawing on ideas from Kant, Hegel, Fuller, Rawls, and so on, in order to flesh out normative arguments: G Fletcher, *The Grammar of Criminal Law: American, Comparative and International, Vol 1* (OUP, 2007).

[22] Again, examples abound, but illustrations would include the direct engagement with what can fairly be expected of a person under duress in the *Erdemović* case, or reflection on fair notice to the individual in the Nuremberg Judgment, or debates about the limits of personal culpability in command responsibility in the *Bemba* case. See eg. *Erdemović* Appeals Judgement, above, Dissenting Opinion of Judge Cassese, para 47-48; Judgment of the International Military Tribunal (Nuremberg), reproduced in (1947) 41 AJIL (supplement) 172 (arguing inter alia that the injustice of prosecution would be outweighed by the injustice of non-prosecution, and that the accused did have a form of notice of the illegality of their acts); *Prosecutor v Jean-Pierre Bemba Gombo,* Judgment Pursuant to Article 74 of the Statute, ICC T.Ch, ICC-01/05-01/08, 21 March 2016, in which each judge gave subtle analyses of the requirements of culpability (see further discussion below, Chapter 7).

the lack of positivity.[23] It can be difficult to show that any particular moral theory or philosophical argument is legally germane. And, of course, the enormous problem with many discussions of moral principles is the lack of concreteness. Further reducing the concreteness, there are many different theories and sets of values, and a lack of guidance as to which is 'correct' or at least authoritative.

Here we encounter a problem that is greater than the positivity-normativity tension. Suppose that we decide not to worry about legal 'positivity' at all: we wish to have a *purely normative* discussion. In other words, we want to discuss what the principles *ought* to be. We might decide, for the reasons discussed in Chapter 3, that we want to avoid assumptions based on formulations of principles in national systems, because they might be inapposite.[24] In this hypothesized conversation, we presumably want to be 'rigorous', and we might understand rigour in the traditional Cartesian way: that we should ground our conclusions in solid foundations. There are two major problems with that conception of rigour.

The problem of insufficient specification

When ICL scholars speak of the moral underpinnings of fundamental principles, the most frequently invoked underlying moral theory is that of Immanuel Kant.[25] Thus, our first thought might be to adopt that as our foundation: we will apply a Kantian analysis to assess whether new articulations of principles are justifiable in abnormal contexts. An attraction of Kant's deontological theory is that it purports to offer an objective, formal, rational, framework which is not dependent on empirical social,

[23] I use the term 'source of reference' to avoid any misunderstanding that I am suggesting that philosophical works are a formal source of law. I am saying, rather, that jurists and scholars routinely (and rightly) engage directly in moral reasoning when making arguments about what is entailed by fundamental principles.

[24] Chapter 3.

[25] I Kant, *Groundwork of the Metaphysics of Morals,* trans M Gregor (CUP, 1998)('*Groundwork*'); I Kant, *The Metaphysics of Morals,* trans M Gregor (CUP, 1996)('*Metaphysics*'). Some features of Kantian thought are that: we must act in accordance with maxims that can be willed as universal law (Kant, *Groundwork* at 15, 31 4:402, 4:421); we must treat individuals as ends and not solely as means (Kant, *Groundwork* at 37-38, 41 4:428-429, 4:433); therefore we can only punish where there is desert (Kant, *Metaphysics* 105 6:331); lawful external coercion is right where it is a response to hindrances of freedom (Kant, *Metaphysics* at 25 6:232); the system is reciprocal coercion in accordance with the universal freedom of everyone (Kant, *Metaphysics* at 26 6:232).

anthropological or cultural inputs.[26] This would be wonderful, as it would sidestep concerns and objections about social contingency or cultural imposition[27]: principles would be derived by logic from a priori premises applicable to all rational beings.

Alas, however, as we look at the conclusions reached by Kant under his methodology of pure reason, we notice that he happens to deduce many of the social institutions familiar in Germany in the 1700s, some of which we would today consider unjust.[28] It seems improbable that those arrangements were dictated by pure reason alone. I am not engaging here in the easy sport of criticizing historical figures for holding views typical of their era. Rather, I am showing the problem with Kant's claim that his theory was not based on empirical inputs (anthropological, sociological, cultural), and hence its promise of neutral rational objectivity. The fact that Kant happened to deduce familiar features of his own society strongly suggests that the process is *not* one of logically-necessary deductions from a priori axioms. Instead, there seems to be considerable gap-filling in deciding what is or is not a 'contradiction', and that gap-filling repeatedly draws on empirical presuppositions and contemporary normative opinions.

Of course, there is nothing wrong with the fact that moral reasoning depends on empirical presuppositions and will be influenced by contemporary values. I am simply pointing out the implausibility of the promise of apolitical, objective, logical deductions from a priori premises. A neutral, objective, system based on rational deductions from universally applicable premises would be wonderful for ICL, because it would avoid criticisms about culture and politics; unfortunately such a system is not available. Even the most ostensibly formal methodology appears to leave a vast latitude as to how one colours in the details.

The general formulas in moral foundational theories usually will not be granular enough to answer the comparatively narrow questions we will be asking about criminal

---

[26] Kant, *Groundwork* at 1-3 (4:388-89)('a pure moral philosophy, completely cleansed of anything that may be only empirical'; 'does not borrow the least thing from acquaintance with [human beings] (from anthropology)', 20-23 (4:408-4:411) (not dependent on 'contingent conditions of humanity'; 'rest only on pure reason independently of all experience'; 'principles are to be found altogether a priori, free from anything empirical, solely in pure rational concepts and nowhere else even to the slightest extent'; 'not based on what is peculiar to human nature but must be fixed a priori by themselves'; 'all moral concepts have their seat and origin completely a priori in reason'; 'they cannot be abstracted from any empirical and therefore merely contingent cognitions').

[27] See §3.3.

[28] For example, approval of second class citizens (*Metaphysics* at 92 6:314-15), no right to vote for women (*Metaphysics* at 92 6:314-15) no right to resist even 'unbearable' abuses by ruler (*Metaphysics* at 96-97 6:320), head of state cannot be punished (*Metaphysics* at 104-5 6:331), mandatory death penalty for murder (*Metaphysics* 106 6:333).

law doctrines. For example, even if we agree that the personal culpability principle requires some 'causal contribution' to a crime, we might see multiple plausible formulations for that requirement (e.g. discernible minor impacts versus risk aggravation).[29] Most comprehensive moral theories will not generate 'answers' to questions of that level of granularity.[30] At best, such theories provide us with helpful ways of thinking about issues and bases for debating them. But our problem is even bigger than the lack of granularity in the moral theories; we also have different moral theories.

The problem of pluralism

An even bigger problem is that there are actually multiple plausible moral theories that could conceivably underpin the fundamental principles. Earlier I referred to 'deontic' commitments. I use the term 'deontic' (i.e. relating to a duty) as a succinct contrast to the relatively simplistic consequentialist arguments often seen in ICL (and national criminal law) argumentation.[31] I use the term 'deontic' to refer to principled constraints rooted in duties to the individual, which we would respect even if doing so does not optimize social welfare.

Given this framing, it is entirely understandable that our first thought commonly goes to the most famous deontological theory, that of Kant. But there are many other moral theories that might underlie and explain the fundamental principles. Within the deontological school of thought, we could turn instead to Hegel. According to Hegel, the criminal law repudiates a person's claim to be entitled to coerce others. On his account, a person has a 'right' to be punished, because it recognizes him as a moral and rational

[29] See §6.8.3.

[30] Of course, consequentalist theories may in the abstract purport to offer determinate answers. In theory there would be, for any given question (e.g. how to formulate a principle in ICL), an answer that in fact maximizes the desiderata of that theory (e.g. utility). But in practice the desiderata will never be perfectly measurable, and hence the problem of insufficient specificity and granularity remains.

[31] Utilitarian arguments in criminal law jurisprudence are often fairly simplistic and incomplete, because they focus on only one variable, namely crime prevention. For example, it is often argued in ICL that we need a broader inculpatory rule for general deterrence, to send a strong message, or to close 'loopholes' that would let accused persons 'escape conviction' (see Chapter 2). Of course, a more sophisticated utilitarian account would grapple with other long-term consequences. These would include the negative consequences of over-criminalization, 'chilling effects' on desirable behaviour, other legitimate social ends (e.g. security or military efficacy), or the optimally efficient limits for the reach of ICL. A more sophisticated consequentialist approach would *reduce* many of the divergences from deontological approaches. However, it would not *eliminate* them, because a true utilitarian would still, for example, punish the innocent if it served the greatest good over the long term.

actor.[32] Or, instead of traditional deontological theories, we could turn to *contractualist* theories to generate the basic principles of culpability and legality. A contractualist theory might look for principles that persons would adopt if they were laying down general rules when negotiating in the 'original position', behind a veil of ignorance as to their actual identity and circumstances.[33] Or, we might look for principles that could not be reasonably rejected by persons moved to find principles for regulation of human conduct that others, similarly motivated, could not reasonably reject.[34] Alternatively, one might adopt a *communitarian* theory and yet still share a commitment to these fundamental principles, if one's theory values autonomy and responsibility.[35] It is even possible that one could construct duty-like limits working within a *consequentialist* model. For example, one could conclude that a criminal justice system can only optimize its benefits in the long run if it posits, as a stipulation within the system, that its officials must strictly respect deontic constraints.[36]

Any of these moral foundational theories might underlie the principles of culpability and legality as we know them. In a liminal case, where we need to further clarify a fundamental principle, each theory might generate a different method of analysis and possibly a different answer. Accordingly, our aspiration of being foundationally rigorous is challenged by two problems: the malleability and imprecision within each moral theory and the plurality of plausible moral theories.

4.2.4 We Have No Reliable Foundation

As the foregoing shows, each of the three commonly invoked sources of reference for fundamental principles is inadequate. First, each source lacks in either positivity or normativity or both. Second, even if we decide to set aside positivity and have a purely

---

[32] By coercing others, the person has recognized for himself a law permitting violation of the freedom of another, and thus has authorized application of that law to himself. GWF Hegel, *Elements of the Philosophy of Right,* A Wood, ed, (CUP, 1991) at 126-27 (§ 100).
[33] J Rawls (edited by E Kelly), *Justice as Fairness: A Restatement* (Harvard University Press, 2001) at 14-18.
[34] T M Scanlon, *What We Owe to Each Other* (Harvard University Press, 1998).
[35] N Lacey, *State Punishment* (Routledge, 2002) e.g. at 188.
[36] J Rawls, 'Two Concepts of Rules' (1955) 64 *Philosophical Review* 3. Or, one could argue – given the difficulty of calculating utility in all cases, as well as problems of dangerous precedents and slippery slopes – that second-order 'rules', including 'maxims of justice', should be followed as they generally advance utility, even if they do not do so in a particular case. J S Mill, 'Utilitarianism', in M Lerner, ed, *Essential Works of John Stuart Mill* (Bantam Books, 1961) 189 at 226-248.

normative discussion about how we *ought* to understand the principles, we are still stuck without a reliable, uncontroverted normative foundation.

A common, even classic, scholarly expectation is that normative analyses and prescriptions should be grounded in some convincing ethical theory. This chapter is, in part, a reaction to the common attitude that there is something suspect or incomplete about a normative argument that is not rooted in a comprehensive theory. In academia, it is common for scholars to select and adhere to one or the other of the main traditions (e.g. Kantian, Hegelian, contractarian, rule utilitarian, etc.) and then offer analyses from that tradition. Such approaches can, of course, offer valuable contributions. However, the problem is that any analyses offered from one particular tradition can be rejected as unconvincing or unproven by any interlocutor who rejects that tradition. As a result, the quest for certain grounding (and of having to declare allegiance to a foundational theory) immediately bogs down in an endless preliminary quest to establish which is the 'correct' foundational moral theory.

If I may state explicitly what is implicit in the classic expectation of certain grounding, the resulting methodology would be:

(1) Figure out which moral theory is the correct one;

(2) Extrapolate from that theory to the best principles of justice;

(3) Evaluate ICL using those principles.

Once the implicit expectation is stated explicitly, it can be readily seen that this is not a feasible approach. After some millennia of trying, we have not determined the 'correct' moral theory, and there is good reason to be skeptical that the answer is coming any time soon (or ever).

Yet there must be a way for us to at least *talk* about more practical ethical questions of the middle range, such as the justifiability of ICL doctrines. Surely we do not have to postpone conversation about fundamental principles in ICL until we first identify the ultimately correct moral theory. There are many issues and controversies to discuss here and now, concerning command responsibility, superior orders, aiding and abetting, co-perpetration, and so on. If we want to make our best efforts to ensure that people are being treated fairly, what are we to do?

4.3 FUNDAMENTALS WITHOUT FOUNDATIONS: MID-LEVEL PRINCIPLES AND COHERENTISM

There is a defensible, thoughtful alternative to starting with foundations. A better approach is to 'start in the middle'. [37] I will outline an account here that works provisionally with 'mid-level principles'. That approach falls within a broader 'coherentist' tradition, which sets aside the quest for certainty and for comprehensive foundations, and instead builds models that promote 'coherence' between the available clues. This account can enable valuable normative and analytical inquiry. I will argue that this is the best means of advancing the conversation in fruitful ways.

4.3.1 Mid-Level Principles

The conceptual tool of 'mid-level principles' came to prominence in discussions of ethics,[38] and has been fruitfully applied in legal contexts, such as tort law and intellectual property.[39] Mid-level principles *mediate* between foundational moral theories and a specific body of practice (e.g. legal doctrines). They are 'mid-level', because they are more abstract and general than specific rules and doctrines, and they are more specific and concrete than comprehensive moral theories. They are relatively discrete and accessible

[37] J Coleman, *The Practice of Principle: In Defence of a Pragmatist Approach to Legal Theory* (OUP, 2003) esp at 5-6.

[38] T Beauchamp and J Childress, *Principles of Biomedical Ethics*, 7th edition (OUP, 2012) (the first edition, working with mid-level principles, was published in 1979); M Bayles, 'Mid-Level Principles and Justification' in J R Pennock & J W Chapman (eds), *Justification* (New York University Press, 1986); M Bayles, 'Moral Theory and Application' in J Howie, ed, *Ethical Principles and Practice* (SIU Press, 1987); B Brody, 'Quality of Scholarship in Bioethics' (1990) 15 *Journal of Medicine and Philosophy* 161; Coleman, Practice of Principle, above; S Diekmann, 'Moral Mid-Level Principles in Modeling' (2013) 226 *European Journal of Operational Research* 132.

Indeed, the idea of mid-level principles has even earlier forerunners; it was foreshadowed for example by J S Mill, who noted 'much greater unanimity among thinking persons than might be supposed from their diametric divergence on the great questions of moral metaphysics... [T]hey are more likely to agree in their intermediate principles... than in their first principles.': J S Mill, 'Bentham' in *Utilitarianism and Other Essays,* A Reid, ed (Penguin Random House, 1987) 132 at 170. Interestingly, Kant also noted the value of 'intermediate principles' in helping to enable judgments about what deeper principles require: Kant, 'On a Supposed Right to Lie from Philanthropy' in Kant, *Practical Philosophy*, at 8:430.

[39] K Henley, 'Abstract Principles, Mid-Level Principles and the Rule of Law' (1993)12 *Law and Philosophy* 121; Coleman, *Practice of Principle*, above; R Merges, *Justifying Intellectual Property* (Harvard University Press, 2011); R Merges, 'Foundations and Principles Redux: A Reply to Professor Blankfein-Tabachnik', (2013) 101 *Calif L Rev* 1361. In tort law, Jules Coleman proposes that the immanent mid-level principle is corrective justice (wrongful loss, responsibility, repair). In intellectual property, Merges proposes that the mid-level principles include proportionality, efficiency, public domain, and dignity.

propositions, applying within a field of practice.[40] Mid-level principles are propositions that are arguably embodied in a body of practice (ie. they *analytically* fit) and also *normatively* attractive.[41] Mid-level principles can be supported by multiple foundational theories: people may agree on the mid-level principles even if they have different underlying reasons to do so. I suggest that the culpability principle and the legality principle can fruitfully be analyzed as 'mid-level principles' in this broader sense.[42]

One virtue of mid-level principles is *convergence*.[43] Participants in a field may agree on certain mid-level principles, even if they differ in their deeper underlying philosophical outlooks.[44] Different moral theories may support the mid-level principles for different reasons, but nonetheless overlap in supporting the principles. This convergence is like Rawlsian 'overlapping consensus'[45] or Sunstein's 'incompletely theorized agreements'.[46] Where such convergence exists, one can fruitfully work with mid-level principles without having to isolate the ultimately soundest basis for them. Of course, there may be some difficult liminal cases, where a principle must be further clarified to resolve the case, and where different underlying moral theories may generate different answers, and thus a choice must be made when specifying the principle. Nonetheless, the mid-level principles provide a valuable *starting point.* Moreover, for many problems of the middle range, mid-level principles are sufficient tools for valuable work, without need for recourse to deeper theories.[47]

---

[40] P Tremblay, 'The New Casuistry', (1999) 12 *Georgetown Journal of Legal Ethics* 489 at 503.
[41] Coleman, *Practice of Principle*, above, esp. at 29. This is the same process as Dworkin's search for analytical 'fit' and normative 'value', at least in his earlier works such as R Dworkin, *Law's Empire* (Harvard University Press, 1986). Dworkin's approach is also coherentist (see below § 4.3.3) and see discussion of coherentism and Dworkin in A Amaya, *The Tapestry of Reason: An Inquiry into the Nature of Coherence and its Role in Legal Argument* (Hart, 2015) at 38 & 46.
[42] As noted in the introduction to this chapter, I am not drawing a hierarchy between 'fundamental' and 'mid-level': I am simply adopting the terminology used in two bodies of literature. I use the term 'fundamental principles' because that is the common terminology within criminal law and criminal law theory: the principles are fundamental within the system of criminal law. I use the term 'mid-level principles' because that is the terminology in the relevant ethics literature; they are at a 'mid-level' between comprehensive moral theories and the legal practice. My argument is that fundamental principles can be fruitfully analyzed as an example of mid-level principles.
[43] Henley, 'Mid-Level Principles', above, at 123 uses the terms 'convergence virtues' and 'practical virtues' ('practical virtues' refers to the relative concreteness of mid-level principles).
[44] Merges, 'Foundations and Principles', above, at 1364-1366.
[45] Rawls, *Justice as Fairness*, above, at 32-38.
[46] C Sunstein, 'Incompletely Theorized Agreements' (1995) 108 Harvard Law Review 1733.
[47] Bayles, 'Moral Theory', above, at 112; Merges, *Justifying,* above, at 9. See for example §6.8.3 and Chapter 7, querying the requisite level of foresight and the requisite level of involvement for the culpability principle.

A second virtue of mid-level principles is that they are more *specific* and concrete than general moral theories, and thus offer more *practical guidance* in a particular context.[48] General moral theories may be too abstract to generate ready answers to many specific problems.[49] For example, if we are wondering precisely what degree of causal contribution is required by the culpability principle, we will likely find that foundational reference points, like the categorical imperative or imagining an ideal conversation, do not generate sufficiently specific answers. Mid-level principles enable us to identify morally or legally relevant characteristics and to note specific normative questions both 'more dependably and more quickly' than we could if directly applying a foundational theory.[50]

A third virtue is that mid-level principles enable an inclusive, pluralistic conversation.[51] Mid-level principles can enable us to debate and often resolve certain concrete problems without first having to agree on ultimate questions of morality. A central theme in my thesis is that discussion of principles is a type of *conversation*. Working with mid-level principles can help facilitate that conversation. As Paul Tremblay observes that mid-level principles 'permit conversation through common language and agreement about normative terms'.[52] Similarly, Robert Merges describes mid-level principles as providing 'a shared language consistent with diverse foundational commitments';[53] they allow us to 'play together even if we disagree about the deep wellspring'.[54]

The relationship between moral theories, mid-level principles, and practice is not just a one-directional deductive chain. In other words, it is not simply that the moral theories support the principles and then the principles dictate the correct rules and outcomes. The interplay is more complex. For example, new cases, or seemingly anomalous bodies of practice, might lead us to reconsider, specify or alter our principles. For example, 'consideration of particular cases and policies can lead one to see effects on

---

[48] Henley, 'Mid-Level Principles' above at 123.
[49] Coleman, *Practice of Principle*, above at 5 and 54.
[50] Henley, 'Mid-Level Principles' at 23.
[51] Sunstein, 'Incompletely Theorized', above, at 1746. Sunstein does not use the term 'mid-level principles'; he refers to 'incompletely theorized agreements' and 'low-level principles', but the idea is very much the same as the mid-level principles discussed here: see ibid at 1740.
[52] Tremblay, 'New Casuistry', above, at 504.
[53] Merges, 'Foundations and Principles', above, at 1364-5
[54] Merges, *Justifying,* above, at 11. In this connection, it is interesting that both Kant and Mill recognize the value of 'intermediate' principles: see above.

moral values and principles not adequately taken into account in the [prior] formulations of mid-level principles.'[55] Or, as Jules Coleman notes, we can use the principles to assess and guide the practice, but conversely the practice also *specifies, concretizes and clarifies* the principles, by applying them to new problems.[56]

Earlier I drew a contrast between 'positivity' and 'normativity'. Often, we might find that 'normativity' can be revealed in the 'positive' (i.e. the practice). In other words, patterns of practice worked out by actors seeking to do justice may reveal plausible implicit underlying conceptions of justice.[57] Thus, practice is not purely subordinate; practice is not merely the object to be *evaluated* by normative tools. The practice may also provide a *clue* helping us to reflect upon and revise the principles as we have formulated them. Accordingly, where a doctrine departs from our current best theory of the principles, there are two possibilities. In most cases, the analytically elegant and normatively sound conclusion will be that the outlying doctrine is problematic and should be harmonized with the principle. But it is also possible that the outlying practice could provide a normative insight that leads us to revise our understanding of the principles.[58] We would strive to identify which solution provides 'coherence' in the deepest sense, which can be a subtle and difficult question, as I will explain further in § 4.3.2 and § 4.3.3.

As I will develop further in the remaining sections, if we take mid-level principles as a starting point, then we can do various tasks, working 'upwards' or 'downwards',[59] and working analytically or normatively. Analytically, we can strive to articulate principles that provide the best descriptive 'fit', and which may be seen as unifying the practice, or at least helping to systematize or guide the practice. This analytical approach can also be used to identify aberrant doctrines (i.e. doctrines that contradict the

---

[55] Bayles, 'Moral Theory', above, at 111.
[56] Coleman, Practice of Principle, above at 54-58.
[57] Of particular interest would be patterns of practice by actors with different foundational moral beliefs; where those patterns are consistent with a unifying principle, then that principle may reflect an overlapping consensus.
[58] In chapter 7, I will argue that command responsibility is an example of a seemingly anomalous doctrine that, on more careful inspection, reveals a useful insight about justice.
[59] I should offer a terminological clarification on the metaphor of 'up' and 'down'. In much of the MLP literature, ethical theories are described as 'up' above, and particular practices and cases are 'down', with MLPs in between. However, in literature on foundationalism, the imagery is of course that 'foundations' are below us, we 'dig down' to the 'deeper' 'underlying' theories so that our arguments are 'grounded.' For consistency, I am adopting the latter metaphor: thus the 'deeper', 'foundational' and 'underlying' moral theories are linguistically and metaphorically 'downwards', and conversely the doctrine is 'above', on the surface.

principles that appear to be immanent within the system). We can also work normatively, asking whether a particular understanding of a mid-level principle is normatively justified, or which of two candidate formulations is normatively 'better'. In liminal cases, the normative task may require descending into competing underlying moral theories in order to flesh out the principles or to choose between formulations.

My argument is that fundamental principles of justice in ICL (and indeed in criminal law) are most fruitfully approached as 'mid-level principles' as the term is used in the ethics literature. Notice that I am not saying that the *only* mid-level principles in criminal law are fundamental principles of justice (such as the culpability or legality principles); on the contrary there are many other organizing ideas in criminal law that are also best understood as mid-level principles.[60] I focus here on fundamental principles of justice because the aim of this thesis is to explore those principles; however, I believe that coherentism and mid-level principles offer an appropriate method for criminal law theory much more broadly.

4.3.2 A Coherentist Account

The proposed approach, of working with fundamental principles as 'mid-level principles', employs a 'coherentist' method.[61] In this section, I will explain the broader method of coherentism. To prevent mis-reading, I should make clear: working with 'mid-level principles' falls within the broader tradition of coherentism, but that does not mean that all of coherentism works with mid-level principles. Thus, when I say that science uses a coherentist method, that does not mean that science works with 'mid-level principles'. Coherentism is the broader category.

---

[60] For example, the 'control theory', or any other theory for delineating between principals and accessories is not a 'fundamental principle of justice' but it is an important postulated organizing concept in ICL. Mid-level principles can be postulated at different levels of granularity and scope. In my view, works exploring the organizing concepts of ICL (or criminal law) are best understood as working with mid-level principles and a coherentist approach. For an example of such a methodology see JD Ohlin 'Second-Order Linking Principles: Combining Vertical and Horizontal Modes of Liability' (2012) 25 LJIL 771.

[61] Beauchamp and Childress, early advocates of the MLP approach, describe their approach as 'coherentist' (pp. 13-25 and 383-385, or see pp. 20-28 in the 6th ed). Coleman, adopting the mid-level principles approach in law, adopts a Pragmatist method, which falls within the coherentist tradition. Coleman, *Practice of Principle,* above, at eg. 6-8.

Coherentism seeks to advance understanding by reconciling all of the available clues as best as one can, without demanding demonstration of ultimate bedrock justification or comprehensive first-order theory. Indeed, the expectation that every proposition should be 'grounded' in an even deeper theory is ultimately unattainable and hence unsound.[62]

Coherentism is the main rival to foundationalism. Foundationalism is the more traditional understanding of justification, in which each of our beliefs should be supported by a more basic belief below (eventually reaching down, ideally, to a reliable bedrock or at least to axioms that are unquestioned).[63] A common metaphor for foundationalism is that the structure of justification is like a building: each floor relies on the floor below for support, until one reaches the foundation.[64]

Coherentism accepts that 'foundations' are not available. Our beliefs do not have to be, and cannot be, rooted in secure or comprehensive foundations. Instead, all we can do is develop models that best reconcile our beliefs and observations; we are given no guarantees of correctness. As William James has written, our beliefs 'lean on each other, but the whole of them, if such whole there be, leans on nothing.'[65] Where a new, inconsistent experience or observation arises, we modify our beliefs to try to reconcile them in coherent schema. We work with all of the available clues, to make them fit as best we can in a coherent understanding.

The foundationalist objection is that such a process sounds problematically circular: belief A supports belief B, and belief B supports belief A. However, that objection

---

[62] 'If anyone really believes that the worth of a theory is dependent on the worth of its philosophical grounding then they would be dubious about physics and many other things.' R Rorty, *Consequences of Pragmatism: Essays: 1972 – 1980* (University of Minnesota Press, 1982) at 168.

[63] The classical foundational would be Descartes' effort in his *Meditations* to derive a set of beliefs from self-evident axioms. Some more contemporary foundationalist accounts are more moderate in that they only require basic beliefs to be 'prima facie' justified but defeasible. However, if such accounts allow a network of considerations to defeat a prima facie assumption, then it seems to me that the method is in the end a coherentist one.

Similarly, some foundationalist accounts could assert that their foundations are simply stipulated as an axiom. The challenge then arises when that axiom is plausibly questioned. Again, I think that the resulting conversation (as we debate the axioms themselves) has to be a coherentist one.

Moreover, any theory that acknowledges that it is rooted in stipulated premises can then be helpfully seen as a simple exploration of what might flow from a certain way of looking at things. For example, what flows if we start from a premise of securing equal freedom? What flows if we start from a premise of maximizing human flourishing? Those are perfectly interesting questions, that a coherentist method can draw upon as valuable *tools*, without accepting any such approach as the ultimately correct and conclusive framework.

[64] See e.g. R Fanselow, 'Self-Evidence and Disagreement in Ethics' (2011) 5 *Journal of Ethics & Social Philosophy*; Amaya, *Tapestry of Reason,* above, at 138.

[65] W James, *Pragmatism* (originally published by Longman Green & Co, 1907) at 113.

itself assumes a linear chain of justification.[66] Instead of the metaphor of a building, the coherentist metaphor is of a web.[67] The coherentist approach is not linear but *holistic:* it aims to refine a *system* of beliefs, rooted in observations and experiences. Our confidence increases the more that our beliefs reconcile experiences and inputs. For many, this approach of reconciling available clues and simply accepting foundational uncertainty may sound disturbingly insecure or even flimsy. I address three main objections (conservatism, uncertainty, and untidiness) below in § 4.3.3.

Perhaps it will provide comfort to recall that the coherentist method matches the scientific method: we form models, we make new observations, and we revise models to better reconcile all the available clues. For example, we can collect diverse clues from fossils, carbon dating, DNA of descendants, and geology, in order to improve our theories about the histories of species and their migration. Each clue in isolation should be approached with caution and skepticism, but we formulate models that best bring the available evidence into coherence, and our confidence in each clue and supposition is bolstered by its coherence with other clues.

One might object that morality is different from science: in science, there can be observations that clearly contradict a model, whereas morality involves more subjective appreciations. However, the methodological similarity is that we still draw on all the *clues* we can. We draw on our analytical application of theories, our intuitive reactions to concrete applications, and even the views and arguments of others, in order to test our ideas and to formulate the best understanding that we can with the available inputs.

Coherentism underlies not only the scientific method but also some normative theories. Examples include the philosophical tradition of pragmatism,[68] the Rawlsian method of reflective equilibrium,[69] and Dworkin's 'law as integrity'.[70] I will therefore cite scholars in each of these traditions for their insights concerning coherentism in general.

---

[66] L BonJour, 'The Coherence Theory of Empirical Knowledge' (1976) 30 *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 281 at 282-286; Fanselow, 'Self-Evidence', above; Amaya, *Tapestry of Reason*, above, at 145 & 535.
[67] Fanselow, 'Self-Evidence', above.
[68] See e.g. James, *Pragmatism,* above*;* J Dewey, *The Quest for Certainty: A Study of the Relation of Knowledge and Action* (Putnam, 1929); Rorty, *Consequences,* above; M Dickstein, ed, *The Revival of Pragmatism: New Essays on Social Thought, Law, and Culture* (Duke University Press, 1998); C Misak, *The American Pragmatists* (OUP, 2013).
[69] Rawls, *Justice as Fairness* above, at 29-32. Rawls' approach is coherentist; for example, he writes: 'A conception of justice cannot be deduced from self-evident premises or conditions on principles: instead, its justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view'. J Rawls, *A Theory of Justice* (OUP, 1999) at 19.
[70] Dworkin, *Law's Empire*, above, esp. at 225-275.

The coherentist approach is *anti-Cartesian* and *fallibilist*, meaning that it does not promise 'certainty'; it openly acknowledges that its conclusions are fallible.[71] Propositions (e.g. mid-level principles) are continually revisable based on new experiences and new arguments. At each juncture, we are formulating the best hypotheses we can to reconcile the available clues. Coherentism is a form of p*ractical reasoning.* It does not strive to unearth the ultimate moral truths; it aims to address concrete human problems and questions as best we can.

My proposed account readily acknowledges that the fundamental principles of criminal justice are *human constructs.* As William James noted, 'you cannot weed out the human contribution';[72] 'the trail of the human serpent is thus over everything'.[73] My account is *post-post-modern*: we acknowledge that each of these principles can be endlessly deconstructed but, rather than falling into nihilism, we are willing to provisionally work with the constructs. We are prepared to question the concepts, but we do not simply discard them all at the outset. After all, if we want to engage in ethical deliberation, then we need to start *somewhere.* We might as well start with the products of the human conversation to date. As Ronald Dworkin acknowledges, 'justice…has a history', and each re-interpretation of it 'built on the rearrangements of practice and attitudes achieved by the last'.[74] Thus, we can take available formulations of principles (for example, that the culpability principle requires that the accused participated in or facilitated a crime in order to be a party to it) as starting hypotheses. From there, we proceed with appropriate skepticism, ready to examine our biases and the historic contingency of current formulations. Thus we are prepared to argue for alterations to existing principles and ideas based on the best available arguments.

A core theme of this chapter is that analysis of principles of justice is not a set of moral deductions, applying some 'ultimate ethical algorithm'.[75] The coherentist accepts that we will not develop a mechanical procedure that can generate correct ethical answers to complex questions; there will always be an element of judgment, and hence

---

[71] Dewey, *Quest for Certainty*, above.
[72] James, *Pragmatism*, above, at 110. He continues, 'Our nouns and adjectives are all humanized heirlooms, and in the theories we build them into, the inner order and arrangement is wholly dictated by human considerations, intellectual consistency being chief among them.' (110).
[73] James, *Pragmatism*, above at 30.
[74] Dworkin, *Law's Empire*, above, at 73-74.
[75] Tremblay, 'New Casuistry', above, at 504 (the quest for the 'ultimate ethical algorithm').

we need deliberation and conversation.[76] We work with human-created, fallible ideas and we do so with human-created, fallible processes. But that is the best and only process to try to discuss the normative justifiability of practices, laws and institutions. Thus, we should embrace the contingency, fallibility and humanity of the conversation. As Richard Rorty has argued, 'to accept the contingency of starting points is to accept our inheritance from, and our conversation with, our fellow humans as our only source of guidance.'[77] Reflecting these themes of humanity and fallibility, Rorty argues:

> Since Kant, philosophers hope to find the a priori structure of any possible inquiry or language or form of social life. If we give up this hope, we shall lose what Nietzsche called 'metaphysical comfort', but we may gain a renewed sense of community. ... Our glory is in our participation in fallible and transitory human projects, not in our obedience to permanent nonhuman constraints. [78]

My account is 'non-foundational', by which I simply mean I am not *relying* on any particular foundation. The account could perhaps even be described as foundationally 'pluralist'[79]: participants in the conversation can draw plausible arguments from different moral theories where they appear to be illuminating, even if we do not yet have a meta-theory that explains how those theories are ultimately tied together. Science does precisely the same; employing models in contexts where they are helpful, even if there are conflicts with other models that work in other contexts, until such time as better models emerge.[80]

My account is melioristic, meaning that I believe that we can *improve* our institutions, practices, doctrines and even our formulations of principles through thought and effort. While the principles may be human constructs, we can still strive to develop *better* human constructs. 'Better' is, of course, neither a simple nor certain matter to assess. 'Better' formulations are ones that better reconcile all the available clues and

---

[76] Rorty, Consequences of Pragmatism, above, at 164.
[77] Rorty, *Consequences of Pragmatism,* above, at 166.
[78] Rorty, *Consequences of Pragmatism,* above at 166.
[79] Merges, *Justifying,* above.
[80] For example, utilitarian and deontological theories both seem to offer valuable insights into particular problems. It is sometimes argued that, because they are seemingly contrasting theories, it is untenable to invoke them both, without at least providing a unifying meta-theory. However, I do not think this is necessarily a problem. For example, suppose science offers two seemingly rival theories: that electrons are waves and that electrons are particles. Each theory is good for handling some problems, but poor for handling others. No one would chide a scientist for invoking each model to handle the problems they are suited for, even if she did not yet have a unifying meta-theory. On a coherentist model, she can justifiably proceed on the grounds, based on all the available clues, that both models seem to be valuable, and that there probably is a good unifying meta-theory even if it has not yet been articulated.

inputs, for example, by more elegantly reflecting the best understanding of what appear to be the underlying values.

In the proposed method, we can take existing mid-level principles (e.g culpability, legality) as provisional starting points. We can then further specify, adjust, or even add or remove principles, based on the best available arguments and inputs. Those inputs include moral theories, patterns of practice, and considered judgments (casuistically testing our sense of justice of the outcomes in particular cases, including hypotheticals.[81]) We seek 'reflective equilibrium': we move back and forth among formulations of principles and our considered judgments of their outcomes in particular cases, adjusting our constructs or re-evaluating our judgments, to reconcile them as far as possible.[82] We look for *deductive* coherence (whether formulated principles match with judgments in particular cases) and *analogical* coherence (whether judgments fit with judgments in analogous cases). More profoundly, we look for *deliberative* coherence, i.e. whether formulated principles cohere with the plausible accounts of the underlying values and goals of the system.[83] Indeed, the enterprise of law itself may entail recognizing persons as agents, and thus we would seek coherence with 'the inner morality of law'.[84] The coherentist method also seeks *elegance* and consilience. For example, a simple principle that convincingly explains multiple features of legal practice offers more explanatory coherence than a series of ad hoc stipulations.[85]

4.3.3 Possible Objections and Clarifications

In this section, I discuss the most important objections to coherentism, namely: (a) conservativism, (b) fallibility, and (c) untidiness.[86]

---

[81] See Tremblay, 'New Casuistry', above. Markus Dubber explores the 'sense of justice', arguing that it involves empathic role-taking with others as fellow moral persons: M D Dubber, *The Sense of Justice: Empathy in Law and Punishment* (Universal Law Publishing, 2006).

[82] Rawls, *Justice as Fairness*, above, at 29-32.

[83] Proposals that improve the coherence between the constraints of a system and its aims have an increased plausibility: see e.g. J Gardner, 'Introduction' to Hart, *Punishment and Responsibility,* 2nd ed (OUP, 2008) at xii-xxxi (constraints and aims of punishment).

[84] L Fuller, *The Morality of Law* (Yale University Press, 1964); K Rundle, *Forms Liberate: Reclaiming the Jurisprudence of Lon L Fuller* (OUP, 2012).

[85] Amaya, *Tapestry of Reason*, above, at 394-96.

[86] There are many other possible objections; in the interests of space I am canvassing the strongest and most salient ones. For much more detailed analysis see Amaya, *Tapestry of Reason*, above, at 57-73, 143-44, 178-87, 308-310, 370-72, 410-412 and 532 (including as to whether coherence is truth-conducive and circularity objections).

(a) Conservativism: The most common initial objection to this method is that it sounds like it cannot be radical. After all, if we work with practice then we are just going to replicate the practice. However, this reaction under-estimates the ambitiousness of coherentism. Coherence is not mere superficial consistency. As I mentioned above, in a coherentist method of identifying the deontic principles, we draw on *all available clues*. We look at patterns of practice for clues about underlying insights of justice, which can include comparative analysis (looking at other jurisdictions, other areas of law, or possibly even other social practices). We look at normative arguments and practical reason, as well as intuition and considered judgments in casuistic testing. We seek coherence in the deepest sense with what appear to be the best understandings of the underlying values.[87] Thus, the coherentist method does not just replicate existing practice.

As Rorty argues, the holistic process of reconciling clues 'often does require us to change radically our views on particular subjects.'[88] Similarly, Dworkin responds to the conservativism objection by arguing that once 'we grasp the difference between [coherence] and narrow consistency' we may come to see that coherence 'is a more dynamic and radical standard than it first seemed', as it encourages us to be wide-ranging and imaginative in the search for deep coherence.[89]

I submit that that the coherentist process of testing incompatible beliefs and practices has engendered the numerous radical changes in human history.[90] Consider for example the abolition of slavery. Slavery was not abolished because someone proved its unsoundness through analytical deduction from an abstract construct, such as the

---

[87] As noted above, there may be values and constraints implicit in the enterprise of law itself; for example, if law is predicated on treating individuals as responsible agents then its doctrines and principles should reflect that L Fuller, *Morality of Law*, above; K Rundle, *Forms,* above.

[88] Rorty, *Consequences of Pragmatism*, above at 168. See also M Sullivan and D J Solove, 'Radical Pragmatism' in A Malachowski, ed, *The Cambridge Companion to Pragmatism* (CUP, 2013) arguing that, although the pragmatist method is often perceived as 'banal', it can be radical in critically assessing both means and ends.

[89] Dworkin, *Law's* Empire, above, at 220. (Dworkin uses the term 'integrity', rather than 'coherence', but the term 'integrity' refers to legal and moral coherence: ibid at 176.)

[90] In this argument, I posit that most people's moral reasoning *is* coherentist. I am making a descriptive claim about how I believe most people in fact engage in moral reasoning. Namely, I do not believe that most people start with a particular foundational theory and then deduce correct actions from it. I think people work with a mass of principles, articulated at different levels of generality or specificity, taking the readily available inputs (e.g. considered judgments), and working in a manner akin to reflective equilibrium.

Aristotlean conception of equality. (Indeed, Aristotle in his formal model carved out an exception for slaves, due to their 'slave nature'.[91]) Instead, people in slave-owning societies reached conclusions and changed their minds based on a wide range of clues and inputs, including diverse important ideas such as freedom, dignity, equality, happiness, as well as empathic responses to suffering. Arguments of the era came to realize that the attempted justifications of slavery entailed jarring inconsistencies with a great many moral beliefs, and hinged on fallacies and unconvincing rationalizations. Empathy assisted these conclusions, as people in slave-owning societies came to grasp that it was an abhorrent and cruel practice. Importantly, even people who had been initially conditioned to accept the practice as 'normal' came to change their mind through this process of reflection and argumentation.

Notice the difference here between (a) broad coherence and (b) superficial consistency amongst a limited set of propositions. Many slave-owning societies (such as in the USA) also espoused principles of equality. At the time, slave-owners argued that slavery was consistent with principles of equality, by arguing that the equality principle applied only between free people and not slaves. At a superficial level, *consistency* between the practice (slavery) and the principle (equality) might indeed be achieved either by abolishing the practice or by declaring a limitation to the principle. But which is the more normatively convincing answer, all things considered? Coherence is a more ambitious and deeper concept than mere consistency amongst a limited set of propositions: coherence requires us to draw widely on all available clues, including ethical arguments, casuistic testing of our judgments based on empathic role-taking, and noticing biases or argumentative fallacies that have previously led us astray.[92] There is

[91] Aristotle, *Politics* (Clarendon Press, 1910) at 1254b 16–21. This is precisely the problem with analytical deductions from abstract constructs: while they purport to be logically pure, they may actually be as distorted and unreliable as any other construct. The best we can do is to constantly test our deductions from any one theory using other theories and judgements (and, iteratively, constantly testing those theories and judgements with other theories and judgements).

[92] The slavery example also shows how 'intuition' can play different roles and how no source of clues is entirely reliable. Some people had the emotional and empathetic reaction that slavery and the cruelties attendant to slavery were clearly wrong. Others had been socialized to see the institution as 'natural', particularly those who directly or indirectly benefited from the institution. Thus, intuition is not necessarily a wellspring of wisdom; it is merely one of the clues taken into account in the search for coherence. Reflective equilibrium calls on us to critically assess even our own intuitive reactions. In the slavery example, we might notice that persons whose intuitions were not disturbed by slavery had all undergone particular social conditioning which was needed to produce that indifference. We might also discover, through analogical testing, that their indifference is severely inconsistent with their reactions to analogous cases. Testing for these types of biases and anomalies would give us reason to doubt those intuitions.

little question that the coherent reconciliation of the full spectrum of available clues is that slavery is wrong.

The example illustrates another merit of coherentism. A proponent of a comprehensive theory rooted in a single value (e.g. freedom, dignity, happiness) might argue that the real problem with slavery was its contradiction of the single value cherished by that theory. But there are many possible values that would entail a rejection of slavery, and many possible theories that could draw on those values with different emphases. Often, we will have vastly more confidence in a mid-level determination (e.g. slavery is wrong) than we have about which supportive theory is the correct one. Coherentism allows us to act on that mid-level determination, even if we do not know which underlying theory is the ultimately correct one.

The process of continually revising our body of beliefs to better reconcile ideas and experiences is even more ambitious than the foregoing suggests, because it is *iterative*: each revision of practices and beliefs in turn enables people to notice, analogically, other practices that conflict with better conceptions of equality. Over time, numerous practices that once seemed natural have gradually been recognized to be discriminatory in various ways. For example, the institution of marriage, which was until recent decades seen as 'inherently' between a man and a woman, has been revised in many societies to include same sex partners and thus to better reflect equality principles. This continual, iterative revision of beliefs and practices is coherentism at work.

My point is that coherentist methods can require radical changes in our beliefs and practices. The continual effort to reconcile our principles, theories, judgments, and practices can lead to the discovery of previously unnoticed latent conflicts. Coherence can therefore require dramatic revision of our beliefs or practices in particular areas.

There is another narrower version of the 'conservatism' objection: since coherentist theories start with pre-existing beliefs (or in this context, widely-recognized principles), they may have a tendency to perpetuate received beliefs.[93] There is merit to this objection. For example, the account I suggest here is willing to accept established formulations of principles as its working hypotheses. By accepting these historically-contingent starting points, there is a risk that one may perpetuate past thinking, and preclude radical thinking.

[93] See e.g. Amaya, *Tapestry of Reason*, above, at 58, 371 & 474.

My response to this narrower objection is that any account has to start *somewhere*. The coherentist account accepts, as a starting point, the conversation that is already underway. It does so because no other more compelling starting point has been identified. If a more compelling starting point were identified, then coherence would require us to start with the new more compelling starting point, and the conversation would shift accordingly.

Consulting formulations developed in national and international practice is valuable as a 'humility check' on our abstract theory-building. One could advance an entirely new normative theory and deduce from it a new set of principles for criminal law systems. However, if we look around the world and notice that no legal system on earth satisfies the proposed requirements, we could rightly take that observation as a clue that there *might* be a problem in the new theory. Patterns of practice, which reflect the understandings of justice of thousands of practitioners over a great many years, are at least a worthwhile checkpoint.[94] Nonetheless, if there are powerful arguments for the new theory, then the totality of available clues might lead us to adopt it. [95]

Earlier (§4.3), I spoke about 'starting in the middle' – meaning starting with mid-level principles rather than with doctrines or with foundations. But there is another way that we must always 'start in the middle': temporally. We start *in media res* – in the middle of the action. The story of criminal law, the story of ICL, and the story of criminal law theory are all already underway. Many doctrines and formulations of principles have already been developed, and certain conversations and debates are underway. So, we start from what has already gone before, we draw lessons and form theories, and we suggest modifications to what is there. Some starting assumptions may later turn out to be 'wrong', but we nonetheless must start somewhere. As in the famous metaphor of Neurath's boat:

> We are like sailors who on the open sea must reconstruct their ship but are never able to start afresh from the bottom. Where a beam is taken away a new one must at once be put there, and for

[94] Parenthetically, another possible objection is that reference to practice seems like a form of the naturalist fallacy – the leap from 'is' to 'ought'. However, as just noted, we consider patterns of practice not to mindlessly replicate them, but out of humility, as they offer clues to understandings of justice worked out by others through extensive practice. Patterns of practice are a helpful common reference point, a valuable check on the imagination of any given individual, and a body of propositions that have at least been tested in practice.

[95] Thus, this is not a Burkean conservative position warning against the unknown dangers of making any changes at all to established social institutions. It simply uses practice as a reference point or possible 'sanity check' in assessing one's own judgements and constructs. Where it is nonetheless clear that practice should be reformed, then it should be reformed.

this the rest of the ship is used as support. In this way, by using the old beams and driftwood the ship can be shaped entirely anew, but only by gradual reconstruction.[96]

(b) Fallibility: The second common objection to the coherentist method is that it does not provide *certainty*. It relies on a series of inputs each of which *might be wrong.* The objection is correct.

The coherentist account freely acknowledges that any of the inputs might be flawed. For example, familiar formulations of principles might replicate biases or blind spots of past legal practitioners. Patterns of practice may be similarly problematic. The major ethical theories are contested human creations and may be gravely flawed. Our sense of justice, or intuition, about particular outcomes might mislead us: it may reflect our prejudices and social conditioning. (There are many possible objections to considering intuition,[97] however, in my view the best argument in favour of consulting our intuition is the outright absurdity of ignoring it.[98]) A coherentist account attempts to reduce error in each of the available imperfect inputs in the only humanly available way: by testing them against all of the other inputs.

An even greater danger still lurks: it is entirely possible that *every one* of those inputs (principles, practice, theories, and judgments) is erroneous. For example, they might all very well be distorted by the same bias, arising perhaps in the human mind or in the human meta-culture. Thus, there is a possibility, not just of error, but of *massive* error.

The coherentist account acknowledges this as well. It acknowledges the fallibility of its inputs and its process, and hence the possibility of error, including potentially

---

[96] O Neurath, 'Anti-Spengler' in M Neurath & R Cohen, eds, *Empiricism and Sociology* (D Reidel Publishing, 1973) 197 at 201.

[97] The strongest objection is that appeals to intuition might just reflect our subjective biases and conditioning and cannot be mistaken for infallible innate wisdom. The objection is of course correct. However, reflective equilibrium does not take intuition as infallible. Both our reasoning (analytical deductions from our constructed theories) or our intuition (reactions to concrete cases) can be wrong. That is why we use each to test the other as best we can. We search for deductive or analogical incoherence and try to inspect both our reasoning and our judgements. This process is obviously fallible, but no infallible process has been identified. The best we can do is test all available inputs against the other inputs.

[98] It seems unthinkable that we would apply cerebrally-constructed moral theories even when our instincts cry out that the results are monstrous. I think such reactions would be a clue that the moral theory might need re-examination. The coherentist approach sensibly uses all available clues. Markus Dubber convincingly argues that the sense of justice requires empathetic identification, and he rejects the dichotomy between emotion and rationality: Dubber, *Sense of Justice,* above at 7-8, 52, 71-72, 83, 146.

massive error. For many scholars and jurists, this acceptance of fallibility is a cause of considerable discomfort. But the response is: *there is no methodology* that furnishes moral certainty. There is no methodology that can guarantee freedom from error or even from massive error.

The expectation of certainty in relation to ethical questions is itself unsound; it is the Cartesian anxiety. The options actually available are either (a) a false pretense of certainty, or (b) a theory that openly acknowledges uncertainty, but which is committed to taking every possible measure for error-correction (maximum corrigibility). The coherentist would argue that the latter is the more mature and honest route. The coherentist abandons 'the neurotic Cartesian quest for certainty'.[99] Instead of seeking certainty, we simply seek to establish *better-justified* principles, drawing on all available arguments. We have uncertain information, uncertain starting points, and we have to make the best decisions that we can with the best evidence and best tools that we can produce. Science proceeds in the same way, often provisionally accepting uncertain hypotheses as starting points to see where they lead, and generating helpful insights as a result.[100]

(c) Untidiness and imprecision: A third set of objections is that the coherentist method is too untidy, imprecise, vague or complex.[101] In other words, coherentism does not provide a clear enough operator's manual on precisely to reconcile inconsistent clues to maximize coherence.

---

[99] Rorty, *Consequences of Pragmatism*, above, at 161. See also JT Kloppenberg, 'Pragmatism: An Old Name For Some New Ways of Thinking?', in M Dickstein, ed, *The Revival of Pragmatism: New Essays on Social Thought, Law, and Culture* (Duke University Press, 1998) (discussing the debilitating 'Cartesian anxiety' that demands 'the grail of objective knowledge' and 'timeless principles'.)

[100] For example, sciences started with a provisional assumption that the reports of our senses map in some way to an external reality. Using the observations of those senses, humans developed theories, made deductions, and built tools that allowed them to learn more and more about the apparent world. We eventually learned that, for example, the world is not composed of 'solid objects' in the way that our senses report; instead 'matter' is overwhelmingly composed of empty space, with fields of energy generating what we perceive as 'solidity'. Similarly, our experience of 'colour' turns out to be a subjective translation of certain forms of radiation. Thus, science shows us some ways that our senses are indeed unreliable. But we got there by using our senses as one set of possibly useful inputs.

If instead, we had said that our senses are not reliable and thus declined to make any further investigations based on them, we would not have worked out that useful information, including about the limits of our senses. In the same manner, a provisional acceptance of familiar formulations, moral theories, and our intuitive responses, provides at least a starting point for deliberations, even if the deliberations may lead us to change our minds about some of those inputs.

[101] Amaya, *Tapestry of Reason*, above, at 57, 143, 181-82.

For example, where an outlying body of practice conflicts with a formulated principle, should we amend the practice, or is the practice a clue (reflecting the practitioners' sense of justice) that should lead to a reformulation or exception in the principle? When there is a conflict between national formulations, moral theories, or considered judgments, which should prevail? The coherentist account does not offer a fixed mechanical protocol for such decisions; there may be plausible arguments for different solutions. Some coherentist thinkers have tried to articulate more precisely what people do when they seek to maximize coherence in their models.[102] Nonetheless, as in science, there is still room to differ about how best to reconcile contradictory clues. For many people, the consideration of so many elements, without a more explicit instruction manual, is too untidy and vague, and thus leaves too much room for individual opinions.

The coherentist response is that it is an unrealistic expectation that a successful theory must provide a clear, mechanical formula that generates morally correct answers. The world is complex. To return to the science analogy, where observations arise that are inconsistent with currently favoured models, scientists often differ on how to reconcile the conflicting clues. Some may adhere to the existing models, with the provisional expectation that the anomalies will be explained away; others may provisionally revise their models, in different ways, to better fit the data. Yet science is not beleaguered with complaints that there should be clear, mechanical rules dictating precisely when a model must be revised and how. Science is not infected with the idea that figuring out really complicated things should be simple.

In a coherentist method, we abandon the 'quest for the ultimate ethical algorithm' that can deductively answer all of our moral queries.[103] Accordingly, reasonable people will at times disagree on how to prioritize and reconcile the clues, just as happens in every other field of inquiry. We can only keep striving to detect unsound arguments, to collect

[102] An illustrative and incomplete list includes: N MacCormack, 'Coherence in Legal Justification' in A Peczenik, L Lindahl, and B van Roermund, eds, *Theory of Legal Science* (Reidel Publishing, 1984); L BonJour, *The Structure of Empirical Knowledge* (CUP, 1985); S Hurley, *Natural Reasons: Personality and Polity* (OUP, 1989); R Alexy and A Peczenik, 'The Concept of Coherence and its Significance for Discursive Rationality' (1990) 3 *Ratio Iuris* 130; M DePaul, *Balance and Refinement: Beyond Coherence Methods of Moral Inquiry* (Routledge, 1993); H Richardson, *Practical Reasoning about Final Ends* (CUP, 1994); P Thagard and K Verbeugt, 'Coherence as Constraint Satisfaction' (1998) 22 *Cognitive Science* 1; K Lehrer, 'Justification, Coherence and Knowledge' 50 *Erkenntnis* 243 (1999); Amaya, *Tapestry of Reason,* above.
[103] Tremblay 'New Casuistry' at 504 (the quest for the 'ultimate ethical algorithm').

more inputs, and to develop better understandings. We cannot eliminate judgment and deliberation from moral reasoning.[104]

(d) Conclusion: In conclusion, all three objections are *correct*: the coherentist approach does not guarantee certainty, it does not provide a precise operator's manual, and it draws on past thought and therefore might perpetuate old assumptions. However, these objections can be made against *any* approach to articulating fundamental principles. There is no method that is certain, straightforward, is divorced from past thought.

These three objections (fallibility, untidiness, and contingency) are actually objections to the human condition. We have imperfect information and no definitive guidance, and the best we can do is to do the best we can do. Since we must build our structures on sand (i.e. fundamental non-certainty), we might as well acknowledge that we are doing so, and attempt to build structures that are as useful and reliable as possible, while also trying to learn more about the sand.[105]

4.4 JUSTICE: A COHERENTIST APPROACH

In this final section, I outline some features of the envisaged coherentist conversation about the principles of justice. (To re-iterate, by 'conversation' I do not mean any special or hidden meaning of the word.[106] I am simply emphasizing that

---

[104] Rorty, *Consequences of Pragmatism*, above, at 164 argues against the Platonic idea that we can substitute 'method' for 'deliberation'.

[105] For those who remain uncomfortable with the coherentist method, regarding it as suspect, incomplete or unreliable, I can offer the following additional responses.

(a) For those who prize 'certainty', I agree that moral certainty would be *better,* but it is not available. If anyone demonstrates the 'correct' moral theory, then I for one would happily root my arguments in the proven-correct moral theory. Until that time, however, we need some other approach.

(b) For those who prize 'reliability', it is arguable that a foundationally pluralist account provides *more* reliability. Where multiple foundational theories could converge in supporting a mid-level principle, we should have *more* confidence in the principle than we would in any one theory. Because mid-level principles are formulated in particular contexts and with comparative precision, some people may support the principle even without knowing precisely which theory they favour in support. As Sunstein argues, for fallible human beings, caution and humility about theoretical claims are appropriate, at least when multiple theories can lead in the same direction. Sunstein, 'Incompletely Theorized Agreements', above, at 1769.

[106] I can say however that process – including inquiry and deliberation – is important in a coherentist account. The coherentist accepts that certain and ultimate truth may be unattainable, and hence I am attracted to the view that the 'best' understanding is that which would be arrived at in an ideal conversation with all attainable information on hand and with all arguments properly considered. This is

refining our understandings is not a matter of mechanically applying an ethical proof; instead it is a process we engage in as fallible humans, using fallible human processes, testing fallible human ideas, and building on the fallible human conversation that has gone before. Different participants with different experiences might draw on different inputs with different emphases; that interaction and debate can provide more clues and more inspiration about the best way forward. Conversations of this nature commonly feature disagreements, but they also feature points that become largely accepted, until such time as those accepted points become disrupted by new insights and better arguments.)

(1) Analytical and normative. First, we can use mid-level principles to work both *analytically* and *normatively*, and to work at different levels of abstraction or concreteness. Analytically, we can try to discern principles immanent within the practice and we can identify doctrines that contradict the best understandings of the principles. Normatively, we can evaluate the competing formulations of principles, we can criticize problematic practices or even criticize principles, and we can try to clarify the best justificatory bases for doctrines. We can work more at the concrete end of the spectrum (assessing doctrines in light of accepted principles) or more at the abstract end of the spectrum (re-examining the principles). The particular emphasis of any given work will depend on the type of contribution it seeks to make: for example, explanatory, justificatory, critical, or reconstructive.

(2) External and internal. Second, a conversation about fundamental principles of justice can adopt a perspective that is both *external* and *internal* to the field of ICL. What I mean is that fundamental principles are both an external normative yardstick by which to judge the system, but they are also internally recognized by the system as interpretive guides or even imperatives. Thus, if we identify a doctrine that conflicts with the best understanding of a fundamental principle, we can make external or internal kinds of claims, and indeed we can make both at the same time. We could say, from an

---

of course only an in-principle aspiration, as we will never achieve an ideal conversation, but I think it correctly states what coherentism strives for (best possible understandings, not ultimate truths). See eg. J Habermas, *Moral Consciousness and Communicative Action* (MIT Press, 1990, trans C Lenhart and S Weber Nicholson) esp at. 43-115, or T Scanlon, *What We Owe to Each Other* (Harvard University Press, 1998); J Rawls (edited by E Kelly), *Justice as Fairness: A Restatement* (Harvard University Press, 2001) esp at xi; Rorty R, *Consequences of Pragmatism: Essays: 1972 – 1980* (University of Minnesota Press, 1982) esp at 160-166.

external perspective, that the system has failed to meet an important normative standard, and criticize it for this failing. But we could also say, from an internal perspective, that the apparent doctrine conflicts with fundamental principles and thus it is '*incorrect*' and needs to be re-interpreted to conform. Fundamental principles are both external tools for criticism and evaluation and also internal tools for clarification and reform.

(3) No fixed priority. Third, there is no single fixed priority among the three commonly-used sources of reference I mentioned above. Those sources were: articulations of principles in ICL itself, general principles derived from national systems, and normative argumentation. Each can properly be used in developing our views, but none of them are paramount.[107] Indeed, we use each source to better evaluate, specify, and understand the others.[108] Each source has different strengths and each is important for different purposes (analytical, comparative, normative). The emphasis appropriately accorded to each depends on the project. For example, a doctrinal project might accord internal formulations the highest priority, but even that project will be informed by induction from national systems and by normative reflection. For a normative project, the hierarchy might seem to be the reverse, with moral theories being the most important. But simple hierarchies still elude us: for example, scrutiny of national and international practice might reveal insights requiring us to revise our normative theories. Even in a normative account, it is valuable to start with mid-level principles and to consult practice. Doing so helps us stay tethered, with humble awareness that even the foundational moral theories are also human constructs.[109] Thus, the process is necessarily recursive and untidy. I think that the back-and-forth process, oscillating between practices, principles, theories and judgments, looking at analytical 'fit' and advancing normative justification or criticism, is an essential part of a grounded normative theory *about* international criminal law.

---

[107] Similarly, Coleman, *Practice of Principle,* above, at 56 declines to assign fixed 'priority' among the practice, the principles and the theories.

[108] For example, a review of national and international practice might reveal plausible underlying intuitions of justice that lead us to revise our philosophical suppositions. Our philosophical reflections may lead us to discern new patterns in the commonalities of national systems that we had not previously discerned.

[109] Sunstein, 'Incompletely Theorized Agreements', above, at 1762: 'But we might think instead that there is no special magic in theories or abstractions, and that theories are simply the (humanly constructed) means by which people make sense of the judgements that constitute their ethical, legal, and political worlds. The abstract deserves no priority over the particular; neither should be treated as foundational. A (poor or crude) abstract theory may simply be a confused way of trying to make sense of our considered judgements about particular constitutional cases, which may be better than the theory. '

(4) A theory of criminal law theory. Fourth, I think that coherentism is also the best explanation of much of the scholarship and juridical practice that works with fundamental principles. I mentioned at the outset that scholarship and juridical argument tends to draw on the three different sources, even though each is flawed. They often do so without explaining why we can draw on those three sources in what might seem to be a hodge-podge. One might expect that I would go on to declare a more correct methodology, or or unveil a fourth alternative, or at least stipulate a priority among the sources. Instead, however, I have concluded that most criminal law theory, at least in this area, is best explained and best supported as an application of coherentist methods. Scholars are drawing on the available clues to construct the best understanding that they can of the principles. If so, my contribution here is largely to make explicit some of the implicit underpinnings of these efforts. I have articulated some of the groundwork underlying much of the scholarship and juridical discourse. If that is right, then we may continue to use all three sources, but simply do so with greater *consciousness* of the limitations of each source and the limitations of the entire enterprise.

(5) A framework for frameworks. Fifth, I am outlining the general framework for a conversation that can incorporate multiple plausible frameworks. I have tried to frame my remarks generally enough to leave space for different outlooks.[110] The justice conversation is inclusive and pluralist, and descends as needed into ethical theories. The method is non-foundational, but it is still receptive to foundational theories, as ways of framing a question and potentially generating helpful insights. A coherentist conversation can still draw on the main moral theories, not as ultimate truths but as '*models*' (i.e. they can show what the implications would be if one focuses on a given set of values or adopts a given set of premises).[111] Contributors to the conversation may bring insights drawing on very different foundational theories; a foundationally pluralist

[110] For example, when I speak of 'deontic' commitments, I speak in classic terms: a duty to the individual that will be honoured even when it does not maximize the social desiderata of typical consequentialist accounts. But it is also possible to advance arguments from 'moderate' or 'threshold' deontology, which permits overrides in the most extreme circumstances. It is possible that moderate deontology has something fruitful to add to the conversation. At the moment, however, we do not seem to have arrived at any conundrum that requires ICL to make a choice on that question. See e.g. discussion in M S Moore, *Placing Blame: A General Theory of Criminal Law* (OUP, 1997), 719-24; T Nagel, *Mortal Questions* (CUP, 1979), 62-63; S Kagan, *Normative Ethics* (Avalon, 1998), 78-94.

[111] Again, this is the same method as is used in science. The overall method is coherentist, which can entail use of one or more 'models'. Where a model has been successful, we can even use it to generate (provisional) deductions, and a highly successful model will be widely adopted. Nonetheless even the most successful model is still a provisional tool and can be discarded on coherentist grounds, such as when a model offering even better coherence emerges.

conversation is receptive to such arguments. Fortunately, a lot of work can be done with mid-level principles without having to descend into their underpinnings or decide between theories. However, there may be liminal cases where it is necessary to do so. For example, what is it about human beings that requires us to afford them respectful treatment? Scholars refer variously to attributes such as agency, autonomy, dignity, the capacity for reason-directed behavior, personhood, worth and so on. For most criminal law problems, we simply would not need to isolate precisely which attributes generate which obligations. However, it is at least conceivable that some criminal law problem may arise that requires us to specify the relevant attribute with more precision, or to decide between different foundational theories. The conversation descends as needed into ethical theories.

(6) Conversation versus contribution. Sixth, when I map out the numerous possible inputs for a coherentist analysis, that does not mean that any given contribution will have all of those features. For example, it is not feasible that any single contribution will canvass all national systems and all moral theories. I am simply aiming to outline some of the tools and moves that can be usefully employed. Different contributors will bring their different perspectives and expertise to bear on different topics of interest. For example, my own contributions will often draw on English-speaking theorists and common law ideas, because that is my experience and expertise and the best way for me to add value to to the conversation at this stage. But in doing so I will strive to engage with the ideas of others, who bring different literatures and legal traditions, in the hopes of building something together that is non-parochial. It will be important for the broader conversation to continue with diverse inputs from diverse contributors, in order to build more thoughtful, durable, and inclusive understandings of the principles.

(7) Hypotheses not answers. Seventh, and finally, the justice conversation will not produce definitive '*answers*'. At best, it provides *working hypotheses* about fundamental principles. The conversation is nonetheless valuable because it requires us to grapple with questions of justice. The discourse around mass atrocity is often dominated by revulsion and the wish that someone be punished; the justice conversation recalls that we must consider the constraints of justice. Of course, there are very different plausible views once we try to specify the principles. (Is criminal negligence

sufficient for culpability?[112] Is causal contribution required for culpability, and if so what does it mean?[113]) We will never arrive at conclusive 'answers' to these questions. In any human enterprise, the best we can do is to make our best efforts to work out the normative underpinnings and to comply with them.

In conclusion, the justice conversation is a fallible, human conversation working with fallible, human constructs. But that does not make it superficial or meaningless. Many readers will be tempted to reject an approach that does not guarantee that the constructs map on to 'true' justice. But we are faced with three alternatives. (1) The first is for someone to discover and demonstrate the guaranteed correct theory of justice. That has not happened yet, despite centuries of deliberation.[114] (2) The second alternative is to give up. Giving up seems far more bankrupt than *trying* to work with the best available evidence. If we care about morality and justice, then we have to try to discuss our practices and institutions – their aims and constraints and overall justifications and possible improvement. (3) The third, and remaining, alternative is to accept that working with the best available clues is the only practicable moral option we have. The best and only assurance we mortals can have that our constructs map on to something meaningful is that our analytical reasoning and intuitive responses tell us so. Thus, the justice conversation may be fallible, human, contingent, and provisional, but it is nonetheless a vital one.

[112] Chapter 7.
[113] Chapter 6.
[114] If someone does discover it, then the coherentist approach would immediately merge with the foundationalist approach in that area, because working with the best clues and models would obviously entail embracing a 'guaranteed correct' model.