Signal to noise

Assessing the role of intuitions in ethical methodology

Kjartan Vedaa Raa

M.A. Thesis in Philosophy
Department of Philosophy, Classics, History of Art and Ideas
The Faculty of Humanities

Supervised by Associate Professor Caj S. Strandberg

UNIVERSITY OF OSLO

December 2016
Acknowledgements

First, I want to thank my advisor Caj Strandberg, for able guidance and a saintly patience with my verbal excesses and thinking-aloud-when-writing.

Second, my family and all the others that have supported me throughout the hideous mental and emotional rigors that – by constitutive necessity, no doubt – must be entailed by any academic thesis worth writing:

Kathleen Henriksen, Verena Issel, Ole Kristiansen, Hildegunn van der Meij, Gabriel Patås VarSkog, Atle Andreassen Raa, Kornelia Raa Vedaa, Liudmila Grigorievna Rezvaya, Anna Felicia Svensson, and Irene Vedaa.

Ordeals like these do pretty short work of any illusions of self-sufficiency one might have harbored, and it is now – to borrow one of the less-objectionable period idioms of Professor Rashdall – intolerably obvious how much my sanity and well-being depends on the people around me. So thanks again, all of you.
Chapter 1: Problems

1.0.0 – Chapter introduction p.7
1.0.1 – Aims and structure p.8

Section 1.1 – Defining ‘intuition’

1.1.0 – What does the literature say? p.11
1.1.1 – Filling out the definition p.13

Section 1.2 – What is the proper dialectical role of intuitions?

1.2.0 – What type of intuitions are the “data?” p.15
1.2.1 – Reflective equilibrium and the primacy of cases p.16
1.2.2 – Signal to noise p.19
1.2.3 – The Only Game in Town p.22
1.2.4 – Two controversial working hypotheses p.24

Chapter 2: Taking sides

2.0.0 – Chapter introduction p.29

Section 2.1 – Hindsight is 20/20: A few object lessons from history

2.1.0 – The presumed incongruities of Mr. Mill’s proposal p.31
2.1.1 – The intolerably obvious realities of Professor Rashdall’s day p.32
2.1.2 – Lord Devlin’s moving and welcome humility p.33
2.1.3 – Where are we heading with this? p.36

Section 2.2 – Reflections on abstraction

2.2.0 – Apples and oranges p.37

Section 2.3 – Finishing touches before we proceed

2.3.0 – Putting the final touches on our initial case for intuition-minimalism p.41
2.3.0 – The obstacle we have to address next p.42
Chapter 3: Distinctions

Section 3.1 – The intuitions that utilitarians rely on

3.1.0 – What do we mean by “utilitarian” intuitions anyway? p.45

Section 3.2 – The “deontological” intuitions

3.2.0 – How do “deontological” or commonsense moral intuitions differ? p.48
3.2.1 – The ones that keep getting lost in translation p.50
3.2.2 – It’s time to get more systematic p.53
3.2.3 – Analyzing the general differences so far p.56

Section 3.3 – There’s abstract and there’s “abstracter”: the tripartite taxonomy of moral intuitions

3.3.0 – All good things are three p.59
3.3.1 – The tripartite taxonomy p.60
3.3.2 – What now? p.62

Chapter 4 – Evaluations

Section 4.1 – The methodological case for macro-intuitions

4.1.0 – We return to the subject of “translatability” p.64
4.1.1 – Universality and precision/clarity p.65
4.1.2 – “Clear and certain” vs. “clear or certain” p.68
4.1.3 – Consistency p.69
4.1.4 – Connecting the threads p.70

Section 4.2 – The epistemic case against meso-intuitions

4.2.0 – The evidence is as yet too circumstantial p.73
4.2.1 – So what kind of intuitions are Greene and Haidt referring to? p.74
4.2.2 – The telltale amygdala p.76
4.2.3 – Interference p.80
4.2.4 – Different processes with different functions p.84

Section 4.3 – Conclusions and implications
4.3.0 – We return to our initial aims p.88
4.3.1 – Implications p.91

Sources p.94
Chapter 1: Problems

The chapter where I declare my aims, state some initial definitions, outline the methodological context, and present my working hypotheses.

1.0.0 – Chapter introduction

Somewhere in the course of the twentieth century, the term “intuition” came to prominence as the preferred designation for the kind of non-inferentially compelling judgments that commonly serve as the pretheoretical “raw data” of philosophy. It is the habit of philosophers to ground their positions on clear and explicit rationales to the greatest extent possible, but there are limits to how far this aspiration can be fulfilled. Every chain of argument must involve some premises that are accepted simply because they seem acceptable, and intuitions are characterized by precisely that: they are judgments that seem compelling to us even if we cannot point out any explicit justification for why we ought to be thus compelled.

Intuitions pose several problems to ethicists, a handful of which we’ll be grappling with in this text. The first and most fundamental problem is that of determining how much intuitions ought to constrain moral theorizing. That is to say: how important is it for a moral theory to accommodate our intuitions? Is it a very grave fault for a moral theory to yield counterintuitive implications, is it not much of a problem at all, or is it somewhere in between? One can hardly overstate the significance of this problem: we examine our initial preconceptions in the context of theory so as to improve our understanding, but we wouldn’t accept an “improvement” so radical as to seem unrecognizable from where we started. Presumably, the proper balance is somewhere in between – but where?

This fundamental problem connects with an additional pair. First, there are the problems we face if we opt to ascribe a very strong theory-constraining role to intuitions: as we shall see, there are reasons to doubt their epistemic credentials, and there are pragmatic-methodological reasons not to rely on them more than necessary. This text will argue in favor of assigning a rather limited theory-constraining and premise-warranting role accorded to intuitions. But such a position – I will refer to it as “intuition-minimalist” – leads to another problem: that of limiting one’s reliance on intuitions without being guilty of inconsistency. Since we need somewhere to start, it wouldn’t do to disavow the use of intuitions entirely. If we’re going to trust some intuitions while distrusting others, we need to point out a relevant, non-arbitrary distinction that can justify such epistemic favoritism.
1.0.1 – Aims and structure

The aims of this thesis are threefold:

a.) To make a general case for why and how we ought to limit the reliance on intuitions. We are at bottom concerned with the very general methodological question of how much evidentiary weight should be ascribed to our intuitions, and how one may justify treating some intuitions as more trustworthy than others. I will argue in favor of limiting the reliance on intuitions quite a lot, and for treating relative degree of abstraction, generality or context-independence as a valid criterion when sorting the wheat from the chaff. I shall be referring to this position as “minimalist,” and the obverse as “maximalist.”

b.) To connect this general case with the more specific debate surrounding the justifiability of utilitarian intuition-deployment practices. As it would be a tad too ambitious to address the very wide and general question above without narrowing the perspective a little, our primary focus will be on how it manifests more specifically in contemporary methodological disputes between utilitarians and the proponents of rivaling normative systems. As a matter of professional shorthand among contemporary philosophers, there is a tacit understanding that the term “intuition” usually refers to intuitions about particular cases. Since utilitarians tend to display a greater-than-average willingness to “bite the bullet” in the face of counterintuitive case-level implications, the methodological disputes between utilitarians and their less-revisionary foils are often phrased in terms of being for or against the trustworthiness of intuitions. However, the informal practice of equating “intuition” with “case intuition” is motivated by convenience rather than any ontological commitments, and it is typically acknowledged that the term “intuition” also encompasses other non-inferentially compelling judgments than just the ones we make about particular cases. It may be said that we are also having intuitions about more general norms and propositions, about fundamental value premises – like the Principle of Utility – and even “methodological” or “formal” intuitions (Nolan 2016, p.164) like the kind of non-inferential judgments we make when a certain premise or framework just seems too arbitrary, ad hoc, or unparsimonious to be in “philosophical good taste.” And when applying this broader understanding of “intuition,” it can no longer be said that utilitarians disavow the reliance on intuition so much as they are selective, for they do seem to rely on value premises and methodological assumptions that are accepted on account of their innate cogency rather than inferential arguments. Such selectivity calls out for an explicit justification, since it would be arbitrary and inconsistent to reject or downplay the evidentiary value of some but not all intuitions without having a clear conception of the relevant features that set the former class
apart from the latter. Although I think this practice of “epistemic favoritism” can ultimately be defended, I am less sure that it has been done adequately at this point. More to the point, it is sometimes (Singer 2005, p.350) justified by pointing out that the intuitions trusted by utilitarians are a very different type, but – perhaps owing to an impatience engendered by the fairly high prima facie plausibility of this claim – the precise details about how they are different do not seem to be outlined in as precise and explicit terms as one might prefer. Therefore, addressing this problem – I’ll be referring to it as the Only Game in Town objection – will be the primary objective that most of this text is structured around.

c.) To contribute to greater terminological precision. Finally, I pursue a third and very general objective as I set out to bring greater clarity and precision into the general vocabulary around intuitions. People keep referring to their intuitions about this case and that case, but it’s fairly uncommon to see our various intuitions treated as a class that can be surveyed and taxonomized from a more general perspective. When someone says “that’s counterintuitive,” would it be possible to specify precisely what intuition the claim is running counter to? If someone says “theory x has a hard time accommodating our intuitions about y,” would it be possible to be more specific about what our intuitions about y entail? Are intuitions irreducibly singular, or could they be meaningfully examined as general classes? Are there subcategories of intuition that we ought to distinguish between? When we intuit in accordance with a recurring pattern, would it be possible to give it a name and state in explicit terms precisely what we are intuiting? My own attempts at answering these questions will take center stage in Chapter 3: Distinctions, and the distinctions proposed in response will remain with us as a general framework throughout the rest of the thesis. These distinctions will play a central role in my justificatory approach at large, although one may find – hopefully – that their applicability extends past just this particular purpose.

Here in Chapter 1: Problems, we shall look more closely into the concept of intuition and outline the definition we’ll be operating with (1.1), before we familiarize ourselves with the surrounding methodological landscape and the problems we wish to address (1.2).

Chapter 2: Taking sides will present the general case for intuition-minimalism and motivate the choice of classical utilitarianism as its more specific focus (2.0-2.3), before outlining the terminological challenges that lie ahead (2.3.1).

Chapter 3: Distinctions will set out to address the need for more fine-grained distinctions within the general class of moral intuitions, examining the intuitions that utilitarians
can be said to rely on (3.1) and the ones they are commonly criticized for not adequately accommodating (3.2), in order to compare descriptively salient differences between the two classes and outline a distinction on that basis (3.3).

Chapter 4: Evaluations will continue by focusing on the methodologically salient differences between the intuitions from the foregoing chapter (4.1), delving more deeply into the underlying considerations that may explain such differences (4.2.), and finally summing up and presenting the finished case for viewing the more-abstract class as commendable in ways that can justify giving it preferential methodological treatment without running counter to the Only Game in Town objection (4.3).
Section 1.1 – Defining “intuition”

The aim of Chapter 1 is to outline the problems that the succeeding chapters will address, but the first thing we have to do is get a clearer grasp of what we mean by the intuitions that pose these problems. We referred to intuitions as “non-inferentially compelling judgments” in the chapter introduction, but this definition needs to be fleshed out a little before we proceed.

1.1.0 – What does the literature say?

Dictionaries may not have the last word when it comes to defining concepts, but it seems reasonable to give them the first. Turning to the Penguin Dictionary of Philosophy, we find that the third and final definition seems closest to what we are having in mind:

Immediate, unreflected belief ‘that we find ready in ourselves as soon as we begin to reflect’ (Bertrand Russell). Intuitions, in this sense of the word, are simply non-inferential beliefs. But the term is sometimes reserved for non-inferential beliefs which are highly resistant, although not immune, to revision or rejection. This usage, which can be traced back to the 1940s, is now adopted by many contemporary philosophers in the analytical tradition. In this tradition it has also become common to appeal to intuitions as the data against which philosophical theories must ultimately be tested. (Mautner 2005, p.310)

This definition points out a couple of elements that are important to us. First, it refers directly to the practice we wish to problematize: the use of intuitions as the “data against which philosophical theories must ultimately be tested.” We’ll return to this phrase in 1.2.0, so we’ll just leave it hanging for now. Second, it notes that the term “intuition” gets used both in a very broad sense (“simply non-inferential beliefs”) and a more restricted one (the ones that are “highly resistant,” albeit defeasible). This calls out for a specification of how broadly we opt to apply the term. As the “intuitions as data” approach seemed to presuppose the narrower sense, we’ll limit our definition to those intuitions that are found non-trivially compelling.

Now that this is clarified, can we run with Penguin’s definition precisely as it is? Not entirely. We need to make one modification. The dictionary described intuitions as “beliefs,” but it may be better to use the broader and more nonspecific term “judgment” within the particular context of ethics. This finds precedent both among philosophers (Kamm 2009, p.23; Sidgwick 1907, p.97) and psychologists (Gigerenzer 2008, p.9).

We now seem to have arrived at a definition that holds intuitions to be non-inferential judgments that are non-trivially compelling. The next subsection will extrapolate on what we
hold these characteristics to entail, but before we get there we ought to look at another definition of “intuition,” this one given by a psychologist rather than a philosopher.

[A] descriptive generalization made by an objective observer about the rule that appears to generate the judgments of an individual about specific cases. And this can be an accurate description by the outside observer, even if the individual would reject this rule if he were to consider it. (Kahneman 2009, p.72)

Now, that was a very different definition. Whereas the philosophical definitions tend to treat intuitions as a type of mental event with certain properties, this psychological definition focused more on intuitions as the implicit rules that certain judgments were in conformity with. And although such rules may have been generative of particular judgments that seemed intuitive then and there, they needn’t themselves be found intuitive when made explicit and evaluated as such. Kahneman’s viewpoint is not unique in this regard, as we find that other psychologists, like Gigerenzer (2008, p.11) also stress the need for supplementing self-report data with outside observation and analysis of judgment patterns as exhibited if one is to properly understand the heuristics – subconsciously applied rules of thumb – that are central to many psychologists’ conception of moral intuition.

We shall be operating with both of these definitions in this text, seeing as they are not mutually exclusive. The second sense (implicit rule) essentially describes a type of intuition instantiated by its various tokens (mental events with certain properties). That is to say, someone may use the term “intuition” in the token-sense when speaking of a particular mental event he is experiencing – he has an intuition that x or it occurs to him that y seems counterintuitive or so on. But for the bulk of this text, our primary interest is in intuitions as types – that is to say, a category denoting a set of intuition-tokens that are similar to each other in some relevant aspect. In Chapter 3, we shall give names to different intuitions, by which we mean intuitions in the type-sense. Thus, we shall then give the name double effect to one particular pattern of intuiting that is often observed, and the name negative hedonism to another pattern of intuiting, and so on. So when speaking of intuition so-and-so, we shall be speaking of patterns of intuitive judgment. We observe that people tend to judge harmful outcomes less harshly if they are foreseen side effects rather than directly causative of an end, and that people tend to disapprove of the infliction of suffering, and so on. We observe that these patterns are consistent, and – recalling Kahneman’s definition – we note that it is possible to describe such patterns in terms of implicit rules. Such implicit rules are what we speak of when we speak of – and eventually compare – intuitions in the type-sense.
But before we start examining different patterns of intuing, we must take a closer look at what we mean by the concept of intuition in the manifest sense – that is to say, what we mean when we say someone is intuing that x or say that intuitively it seems to me that y or so on.

1.1.1 – Filling out the definition

Intuitions in this sense – the token-sense – are a category of mental event. As mentioned before, we define intuitions as judgments that are experienced as compelling without their holder being consciously aware of any consideration he could refer to if asked why he finds them compelling. That is to say, their justification (if any) is not introspectively accessible, but they come bundled with a second-order propensity to respond to them as if they had one. This makes intuitions different from percepts, memories and inferences, as people have a clear idea of what conditions must obtain if these judgments are to be justified. Perceptual beliefs are justified if they accurately reflect the world, memories if they accurately reflect the past, and inferences if they are drawn from justified premises. In contrast, it is very opaque what kind of condition must obtain for an intuitive judgment to be justified. It just feels as if it must be.

We mentioned that the type of intuition that interests us should be compelling to more than a trivial degree. This postulate is less a matter of ontology than of methodological parameters: it is mainly a question of what sort of intrinsically compelling judgments are interesting for our purposes. Arguably, any kind of judgment can be called somewhat compelling in virtue of presenting itself to one’s consciousness at all. But some judgments are only experienced as compelling to a minimal degree. If someone says he “acted on a whim,” he implies that he was impelled by an action-guiding impulse that was highly transient and only trivially compelling. This makes a whim different from an action-guiding intuition, for intuitions are compelling to a degree that is not transient and not trivial. If someone restrains himself from acting on a whim, the impulse will dissipate. In contrast, intuitions are recalcitrant. If someone is in the habit of acting on whims, people will say that he is uncommonly capricious, impulsive, whimsical, etc. This would not be the case with someone in the habit of acting on his intuitions. They might say he trusts his instincts or goes with his gut, but a habitual compliance with

1 It should be specified, however, that although intuitiveness presupposes intrinsic compellingness or cogency, this cogency can still overlap with other, extrinsic types of warrant as well. For instance, someone who is challenged to justify the premise of pain as a sui generis evil may just as well opt to justify it on another basis than intuition – for instance, the methodological claim that it is apt to use common-ground beliefs as a starting point, combined with the empirical claim that the premise in question qualifies as such – while still perceiving the premise as intuitive at the same time. Generally speaking, philosophers will prefer not to rely on intuitive warrant more than necessary, since playing the intuition card is essentially a way of closing the discussion about the premise in question.
intuition does not carry the connotations of inconstancy that whimsicality does. Likewise, people sometimes find that doxastic notions pop into their consciousness with no discernible origin – this category would include such things as stray associations, flights of fancy, and puzzling notions that seem to occur from nowhere while in a state of dreaming or poetic rapture. Like intuitions, these belief-states are experienced as non-inferential: their origin is not introspectively accessible. But like whims and unlike intuitions, such non-inferential belief states are no more than trivially convincing. Whereas whims are mainly trivial on account of their transience, notions-from-nowhere are more often trivial in the sense that they are unlikely to be ascribed anything but the most paper-thin epistemic warrant on their own. Someone with a high tolerance for equivocation might say they are considered self-evident in the sense that whatever warrant they do have is intrinsic, but not self-evident in the sense of supplying enough warrant to justify much at all. While people sometimes do attach relevance to the content of such “orphaned” notions, this tends to be accompanied either by an apologetic air (“it’s just a weird notion that popped into my head… surely means nothing”) or assumptions about what its “true” source may be (“the contents of this dream may sound meaningless, but I think my subconscious is trying to tell me something important”). In contrast, notions that are non-trivially intuitive tend to be taken far more serious, even if their holder is at a loss to give any account of why that should be so. There is typically a difference in perceived modality: non-inferential but trivial beliefs tend to seem contingent, whereas a notion found intuitive will often seem as if it could be no other way.²

A final point to note is that the boundaries between intuitions and emotions can sometimes be unclear. There is a great deal of entanglement between the two. Some intuitions give rise to emotions, and it may also be that some emotions give rise to intuitions. Intuitions often occur alongside emotions and may be triggered by the same factors. This may be a matter of dual causation, but it’s also a possibility that some intuitions are identical with emotions. It is now common among moral psychologists to treat moral intuitions as closely connected or even identical with emotions, but this stance is not as popular among moral philosophers (Woodward and Allman 2007). I won’t take a stand on the ontological relation between emotions and intuitions, but from a methodological point of view, it is advisable to treat them as two distinct

² There is another boundary to note. If one defines an intuition as a non-trivially compelling judgment with no introspectively accessible origin or grounds of justification, it can be argued that this concept includes the type of judgment that occurs when an entailment relation is being recognized (that is to say: seeing that something “follows” in logic or mathematics). However, I opt not to operate with such a broad definition of intuition as this. Although there is no conceptual incoherence in saying that the process of apprehending entailment is a form of intuition, excluding it is preferable from a methodological point of view. It just isn’t subject to the same kind of doubt and controversy as the kind of judgments one usually refers to when speaking of intuition in philosophy. We are then concerned with judgments that are found non-trivially compelling despite lacking inferential justification, but not compelling to such an overwhelming and uncontentious degree as the apprehension of entailment.
classes. The complicated relation between intuition and emotion is, as Chapter 4 will explore, a matter of considerable significance.

Section 1.2 – What is the proper dialectical role of intuitions?

1.2.0 – What sort of intuitions are the “data?”

We have defined intuitions as intrinsically compelling judgments, and inasmuch as they are to be moral intuitions they ought to be intrinsically compelling judgments about matters that pertain to morality. This is very broad, and seems to encompass at least four different classes of intrinsically compelling judgments.

a.) First, there are judgments about particular cases, like “x ought to do y in situation z,” “the behavior exhibited by x can correctly be described as an instance of y,” or “it is a preferable state of affairs if x rather than y.”

b.) Then there are judgments about the aptness or significance of more general norms or assumptions that a theory can seem implausible for failing to accommodate, but which may not always be found that intuitive when evaluated as theoretically basic propositions.

c.) When some initially-intuitive norms or general assumptions seem less intuitive on closer reflection, this is often because they run afoul of our methodological intuitions. Although not all methodological assumptions rely on intuition as we have defined it (the need to avoid contradictions is so uncontroversial as to fall outside of our definition, for instance), many do. We generally disapprove of postulates that appear too ad hoc, arbitrary, tautologous or explanatorily vacuous, but we seldom have very clear and explicit criteria for when a postulate crosses such a line. For instance, there are some (e.g. Greene 2013, pp.223-224; Scanlon 2008, p.18) who view the Doctrine of Double Effect as too devoid of theoretical justification to be philosophically satisfactory, while others again (e.g. Nagel 1986, pp.179-180) view its theoretical deficiencies as adequately counterbalanced by the value of its conformity with our pretheoretical judgments. Likewise, we often have to rely on our “theoretical gut feeling” when deciding if a certain theoretical framework sacrifices too much parsimony or simplicity for the sake of conforming with our pretheoretical judgments, or making a judgment call about when a theory can be viewed as displaying too little precision, interconnection, internal consistency, conceptual symmetry, or independent plausibility.
d.) Finally, we have intuitions about highly abstract and fundamental notions, such as very generalized and cross-contextual criteria of value or rightness, like the Principle of Utility or Categorical Imperative. When these are classified separately from our intuitions about general-but-not-that-general norms or assumptions (b.) it is because they are less likely to conflict with our methodological intuitions, which tend to place a high value on a premise’s ability to connect many disparate judgments with as much parsimony and as little ad hoc-ness or inconsistency as possible.

Returning to the phrase we left hanging back in 1.1.0, we notice that the dictionary entry did not specify what kind of intuition it was “common to appeal to as the data against which philosophical theories must ultimately be tested.” It certainly makes a big difference whether the prevalent practice in the field is to treat all, some or even just one of the classes as the data against which theories must “ultimately” be tested.

Would it be accurate to describe the prevalent methodological outlook of today as privileging any intuition-category like that? The answer is ambiguous. In one sense, it would not be accurate. This is because the most widespread approach of contemporary ethics – the method called reflective equilibrium – is one that aspires to let all four classes have their due in evenhanded measure. But in another sense it may not be so inaccurate, seeing as there is a general tendency to ascribe case intuitions (a.) with a particular degree of methodological prominence that may be characterized as halfway-analogous to that of observation data in empirical fields.

1.2.1 – Reflective equilibrium and the primacy of cases

One of the main challenges of doing normative ethical theory is that one is seldom more than a hairsbreadth away from category error. There is no choice but to take descriptive statements about the moral judgments we do make – as a matter of fact about our psychology – as the starting point of our investigation, but our ultimate aim is to derive normative conclusions about how we ought to judge. And it’s not entirely clear what kind of requirements must be fulfilled if one is to be justified in using premises from the former category as the basis for reaching conclusions in the latter. Nowadays, the method of reflective equilibrium – we’ll simply abbreviate it as “RE” from now on – is the most widely accepted “recipe” for how this descriptive-normative gulf can be legitimately crossed.

---

3 This is not to say that such principles are intuitions, but there is typically intuition involved when we assess how independently compelling or reasonable they seem to us.
The first stage of the procedure is almost purely descriptive: one surveys the initial, psychological facts both about our general notions (b. and d.), and the judgments we make about various cases (a.) and sets out to formulate a theoretical framework of unifying principles from which these notions and case judgments could be generated. Already here, however, there is a slight normative element involved as there are threshold criteria – like the absence of partiality or similar distorting factors – that a potential input must satisfy lest it be excluded on the outset. So at this point, the facts about our intuitions are treated in the same way as raw data in the empirical sciences: one excludes observations made under particularly unfavorable conditions, but the rest of them are recorded as completely and faithfully as possible. But whereas the empirical sciences operate with a very high threshold for when it is legitimate to discount observations that are in conflict with one’s theory, the RE practitioner has a good deal of leeway here. A moral theory that attempts to mirror all of the input that seems plausible on the outset is unlikely to be found very satisfactory on abstract reflection. Much of the raw data may turn out to be inconsistent, mutually contradictory or otherwise unable to satisfy straightforward (non-intuitive) methodological criteria, while other parts seem to make the resulting theory methodologically unsatisfactory in a less definite, more intuitive sense (c.). Others again are found hard to reconcile or account for with principles that appear plausible or otherwise appealing on abstract reflection (d.). After amending the theory so as to be more satisfactory from the abstract point of view – less inconsistent and with fewer elements that seem ad hoc, more parsimonious and with greater independent plausibility, and so on – the RE practitioner will then shift his focus back to the non-abstract perspective to assess how well this amended framework can accommodate our case judgments and commonsensical notions. If the resulting revisions are a bit much to swallow – perhaps the amended principles lead to outrageous case-level implications, for instance – we may opt to re-amend the offending principles so as to bring their implications more in line with what we are willing to countenance. This, however, necessitates another shift to the abstract perspective, as we ask ourselves how well these modifications mesh with our theoretical sensibilities, perhaps finding that certain counterintuitive case implications may after all be tolerated for the sake of retaining an especially appealing principle, while other case intuitions in turn are so compelling that we are willing to sacrifice some desirable theoretical qualities for the sake of accommodating them (DePaul 2006, pp.598-604; Lazari-Radek and Singer 2014, pp.96-104).

This process – a sort of back-and-forth, give-and-take between what seems most intuitive in the abstract and most intuitive at case level – may then be repeated as many times as necessary
in order to reach a point where the theorist feels satisfied with having found the best possible balance between the competing concerns. Such a balance – a state of harmonious alignment between our often-conflicting intuitions about what seems plausible from the abstract as opposed to the concrete perspective – is what the term *reflective equilibrium* refers to (Lazari-Radek and Singer 2014, pp.96-104).

But even if RE can in this sense be said to take all types of intuitions into account, it may not be that *this* is the most accurate interpretation of the “intuitions as data” practice the dictionary was referring to. This is because, as mentioned in the introduction, the term “intuition” is commonly understood to be shorthand for “case intuition.”

Weinberg (2016, p.296) uses the term *primacy of cases* to describe this common presumption. Although case judgments are seldom held to *exhaust* the range of judgments subsumed under the term intuition, they do tend to receive the lion’s share of attention, and – more so – they often seem to be ascribed greater evidentiary weight than more abstract intuitions. When intuitions are described as “constraining the discourse” (Mallon 2016, p.412) or acting as “fixed points” that moral theories are expected to accommodate (Sunstein 2013, p.2), the intuitions in question are generally case intuitions rather than intuitions about general notions, formal premises or methodological propriety. If it can be demonstrated that a moral theory leads to strongly counterintuitive case-level implications, this is often held to be a very grave defect that cannot be fully counterbalanced by the intuitive appeal of its grounding premises. In such conflicts, there is usually a lower threshold for doubting our premises than for accepting the counterintuitive case implication they lead to (Weinberg, *ibid*).

One may imagine a couple of reasons for this. First, our intuitions about particular cases tend to be stronger or more *salient* than our intuitions about abstract principles. Second, it is often possible to limit the applicability of an abstractly intuitive premise without having to *reject it wholesale*. It may be enough to “downgrade” it to being valid *pro tanto* – “as far as it goes” – by introducing some additional premise that acts as a *side constraint* or *defeater*. In contrast, such compromise solutions are unlikely to be available where concrete cases are concerned: either it *is* or *is not* morally right to do a particular act that that seems intuitively unacceptable. Owing to considerations like these, it may be understandable that it has become quite common to treat case intuitions (a.) as especially vital inputs when practicing the RE approach. So in that sense, it wouldn’t be inaccurate to say that case intuitions are indeed being treated as “the data.”
1.2.2 – Signal to Noise

One might say that the question of which intuitions to regard as “the data” connects closely with one’s assumptions about the overall signal-to-noise ratio of our intuitions. If we assume that our intuitions are for the most part reasonable and well-founded – that is to say, there is much signal and little noise – it makes sense to give all of them the benefit of the doubt and assume that the minority of intuitions that are misleading may be unmasked once we understand the underlying rationales of the majority that are trustworthy. In contrast, someone who suspects that our intuitions’ signal-to-noise-ratio is not so good – that many of them are misleading – would be more likely to view such a project as a misguided search for something that’s not available.

We can conceptualize these two positions as “maximalist” and “minimalist” respectively. While the maximalist approach assumes that the bulk of our case judgments are sensitive to some implicit rhyme and reason that may be uncovered through examination, the minimalist approach suspects that such rhyme and reason must be imposed from without through the judicious application of principles that we find reasonable already on independent reflection. However – as we already mentioned – doing this still involves intuitions about which principles do seem independently plausible (d.), and precisely what it takes to apply them judiciously (c.). Nonetheless – even if the approach cannot avoid intuition entirely – it is still likely to keep the reliance on intuitions as low as it can be, which is desirable to someone who suspects them of being a generally unreliable form of data.

Owing to the considerations presented in the previous subsection, someone with a maximalist approach is likely to be hesitant to contradict or disregard case judgments (a.) when they conflict with notions that – although compelling in the abstract (b. or d.) – may yet be modified so as to accommodate our case judgments, provided this modification is not found too methodologically distasteful (c.) on grounds of adhocness, nonparsimony or the like. In its most extreme form, the epistemic privileging of case intuitions manifests as particularism – i.e., the stance that holds moral judgment to be too irreducibly context-sensitive for general principles to play a helpful role of any significance – but our focus will be limited to the approaches that are generalist. When a generalist methodology favors case intuitions above abstract intuitions, it may be characterized as taking a bottom-up approach (Waal 2013, p.228), in which abstract premises are often derived by generalization from particular case judgments (Hintikka 1999, p.134). Although a bottom-up approach doesn’t have to entail treating a principle as positively justified in virtue of conformity with case judgments alone – for to do this would turn the
ethicist’s project into a purely descriptive one – it is likely to involve a rather high threshold for when it is acceptable to conflict with them. Bottom-up generalists will still expect something more from a moral principle than merely accommodating the relevant set of case judgments (Kamm 2009, p.23). They may, for instance, seek to make more sense of our case judgments’ internal logic by generalizing them in terms that plausibly connect with some overarching, morally significant concept like personhood or rationality, which in turn may help us to uncover inconsistencies or dispel perplexion about unusual or ambiguous cases. Ethicists who favor case judgments will still be sensitive to more general methodological considerations, but they are likely to display a greater willingness to assume that our case judgments are – by and large – sensitive to some underlying rhyme and reason that we would find coherent and plausible if uncovered. This approach is widespread in contemporary ethics, especially as manifested through what is sometimes called the Method of Cases (Baz 2016, pp.115-116).

In contrast with this, someone who favors a minimalist approach is likely to be especially sensitive to the methodological intuitions (c.) that favor those general notions that seem most conducive to theoretical parsimony and consistency of applicability (d.). In contrast, the intuitively-plausible notions that do not satisfy such criteria so well (b.) are often distrusted alongside the case intuitions (a.) that they often can derive support from.

It may not be surprising that the utilitarian theory – characterized by its reformist bent and strident skepticism towards many elements of commonsense morality – is notable for bucking the modern trend towards intuition-maximalism. Whereas nonutilitarians sometimes treat intuitions with a reverence close (but not equal) to that extended to observation data in physics, the utilitarian stance is more akin to that displayed by physicists towards their intuitions about their field of study, which are not just distrusted readily but at times almost as a point of professional pride. Speaking of which, the overall theoretical outlook of utilitarianism does seem to display quite a few similarities to that of physics, seeing as it tends to mirror the ardent reductionism, delight in upending familiar worldviews, and perceived overfondness for mathematics that sets the latter field apart from the special sciences.

In contrast, the opposing approach – the one I have called intuition-maximalist – may be said to display closer parallels to a quite different empirical field, namely linguistics. It has been argued (Hintikka 1999) that the perceived success of Chomsky’s approach in linguistics may have played a large part in giving rise to the modern style of intuition-deployment. Whereas the Penguin dictionary traced the origins of the intuitions-as-data approach back to the 1940s,
Hintikka views the 1960s as closer to the mark, and – seeing as he has more than a few reservations about the practice – describes it in less-than-approving terms as an “attempt to get on the bandwagon of transformational grammar” (p.127). Questions of ultimate motivation aside, the comparison between linguistics and the bottom-up style of intuition-deployment need not bring resentment from its proponents, seeing as similar analogies have found endorsement from within their own ranks (Mikhail 2011). Just as many linguists will judge a proposed grammar’s adequacy by its success in generating those and only those sentences that are judged acceptable by our intuitive Sprachgefühl, many proponents of the Method of Cases seem to place a very high premium on a proposed moral grammar’s ability to generate implications that – by and large – stay within the bounds of what our intuitive Sittengefühl4 finds acceptable. More so, both groups tend to view these intuitive acceptability judgments as research data that can be profitably analyzed in terms of underlying patterns in order to arrive at a better understanding of the introspectively inaccessible capacity – the deep structure – from which they are generated.

Seen through the lens of these analogies, one might say that the signal-to-noise question is one that connects closely with one’s views on the availability of countervailing evidence that can outrank our moral intuitions. Physical intuitions are easily discarded when they conflict with theory because one is confident that the theory is founded upon other, far more trustworthy modes of evidence. In contrast, linguistic intuitions tend to be viewed as intersubjective facts on the ground that a theorist can ill afford to refuse accounting for, seeing as there aren’t any immediately obvious candidates for what kind of evidence should be more trustworthy when investigating this domain. It is unlikely to draw much protest if one hazards that the proper methodology of ethics should be somewhere in between these two extremes, but it is hard to know precisely where.

Trying to model one’s approach too closely on linguistics is problematic on several fronts. For one, our moral judgments do not display anywhere near the uniformity and consistency of grammatical acceptability judgments within a linguistic community, which does seem to obviate a rather major element of what made our linguistic deep structure a subject of interest in the first place. Second, fidelity to our Sprachgefühl is mainly viewed as an important success criterion by descriptive linguists, owing to our grammatical intuitions’ blithe insensitivity to many of the qualities that interest their colleagues in the prescriptive or normative end of the field, such as logical rigor, aesthetic appeal, and communicative

---

4 Yes: that’s not a real word, and yes: I checked with a German to make sure the neologism was grammatically stringent and made semantic sense.
expediency. If our Sittengefühl should turn out to be similar to its linguistic counterpart in this regard – that is to say, lacking in sensitivity to considerations we would recognize as properly normative – this would make it rather unpromising as a source of theoretically useful moral data (Lazari-Radek and Singer 2014, p.105).

But the physical-minimalist approach has challenges of its own. It presupposes that we have some kind of evidence that we can rightly consider trustworthy enough that we can justifiably discount case intuitions in conflict with a theory founded upon it. As it turns out, this evidence will itself have to include intuitions, which leads us to what the introduction referred to as the Only Game in Town objection.

1.2.3 – The Only Game in Town

Since the common practice is to use “intuition” as shorthand for “case-level intuition,” it follows that many of the disputes between utilitarians and their foils will be phrased in terms of being for or against intuition. Thus, many utilitarians will be concerned with undermining the evidentiary status of intuitions, which – as we shall see in the coming chapters – is a fertile field. However, this kind of anti-intuition approach can often be thrown off balance when their deontologist colleagues remind them that an attack on intuition is not just an attack on the intuitions they oppose, but also on their own basic premises, which are themselves intuitions in the wider, less informal sense of the word. To an uncharitable observer, this maneuver may at first seem like a bit of linguistic trickery: simply a strategic switching between the formal and informal senses of a word in order to effect a dubiously warranted tu quoque. But upon closer reflection, the move turns out to be a legitimate objection on grounds of inconsistency. Let’s have a look at how this kind of dispute may manifest in the literature:

I disagree with Singer on methodology. First, he also relies on some intuitive judgments – those about the plausibility of general principles, such as “maximize the good.” I think that our intuitions about general principles must be tested against our intuitions about the implications of general principles for cases. […] I do not think that our intuitions about cases are less reliable than those about principles. This is, in part, because we do not have a reliable grasp of a principle itself without considering its implications (Kamm 2011, pp.417-418).
Since Kamm was referring to Peter Singer, it seems a natural choice to see what opinions he might have voiced on these matters.\(^5\)

It might be said that the response that I have called “more reasoned” is still based on an intuition, for example the intuition that five deaths are worse than one, or more fundamentally, the intuition that it is a bad thing if a person is killed. But if this is an intuition, it is different from the intuitions to which Haidt and Greene refer. […] It may be closer to the truth to say that it is a rational intuition, something like the three “ethical axioms” or “intuitive propositions of real clearness and certainty” to which Henry Sidgwick appeals […] (Singer 2005, pp.350-351).

Kamm suggests that as long as “intuitiveness” describes the type of warrant used both when assessing a principle’s initial plausibility and the plausibility of our judgments about its case-level implications, there is no obvious justification for saying that the latter form of intuitiveness should count for any less than the former. Singer, in turn, suggests that the intuitions underlying utilitarian reasoning are “different from” the kind of intuitions he dismisses. Unlike the latter, they may better be described as “rational intuitions” and perhaps likened to the “ethical axioms” of Sidgwick.

So here’s the core of the Only Game in Town problem: utilitarians tend to distrust case intuitions when they conflict with the implications of their theory. They justify this by giving reasons for holding intuitiveness to be a poor source of warrant. But when the deductions of their own theories are traced all the way back to their foundational premises, they often find that it is difficult to justify these without appealing to their intuitiveness. Thus, it seems intuitions can properly be called the only game in town. Even someone who disavows the reliance on intuition in the informal sense of case judgments will still have to rely on intuitions in the broader sense. If there is no way around using some intuitions, the least arbitrary option is to accept all intuitions without prejudice. If someone wishes to play favorites by privileging his favored premise-intuitions above the case-intuitions he disdains, the onus of justification falls upon him. He must point out some relevant difference between the two types.

To point out such a difference is the goal of the coming chapters. I wish to examine the relevant differences between different types of intuitions in order to find valid reasons for treating them differently. In other words, I wish to argue that it is a false dichotomy when one suggests that the only non-arbitrary options are to either rely on all intuitions or no intuitions.

---

\(^5\) In the interest of not inadvertently misrepresenting anything, it ought to be specified that Singer’s statement is – perforce, seeing as it’s made at an earlier date – not intended as a direct response to the particular Kamm quote reproduced here.
1.2.4 – Two controversial working hypotheses

Throughout this section, I have attempted to present a rough outline of certain aspects of the contemporary normative-ethical methodological landscape, emphasizing the distinction between the minimalist approach pursued by many utilitarians and the more maximalist approach now prevalent in the field at large. I intend to argue in favor of the former approach, largely on the basis that the latter methodology presupposes an unwarranted degree of trust in our moral intuitions as a general class. Justifying this claim will in large part involve examination and evaluation of those intuitions, which will be the central focus of Chapters 3 and 4. Obviously, one should never let one’s investigations be too rigidly constrained by initial expectations of where they might lead, but – on the flip side – it is seldom if ever a realistic aspiration to examine a complex range of phenomena without some initial, structuring assumptions about what to look for. In order to provide such structure and direction, it is necessary to have some initial working hypotheses.

Working hypothesis #1 – Ethicists should distinguish more clearly between normative and descriptive success criteria: The most prevalent outlook of modern ethical methodology – that which I have referred to as linguistic or maximalist – is an approach that, in my humble opinion, suffers from a tendency to conflate descriptive success criteria with normative ones. What do I mean? When comparing the relative merits of differing normative systems, it is common to place a high value on their descriptive success in accommodating our case judgments and pretheoretical assumptions. I will not dispute that such descriptive success criteria are relevant to normative ethics, but I believe the degree to which they are significant has become unduly overemphasized. Just as fidelity to our Sprachgefühl is a far less valuable quality to the normative linguist than it is to his descriptive counterpart, I believe fidelity to our Sittengefühl to be among the comparatively less important qualities that have bearing on a normative moral theory’s merits. The qualities that make a moral theory good as a descriptive account need not be the same as the ones that make it good as a source of normative guidance, but this distinction – or so I will assert at any rate – is not being adequately recognized by the prevalent outlook in contemporary normative ethics.

Working hypothesis #2 – Bentham’s underrated: My first working hypothesis could hardly be called uncontroversial, and the same goes for my second. Put shortly, I believe that when evaluated in accordance with the success criteria that in my opinion should take precedence in normative ethics, the utilitarian theory – and by this term, I will be referring to hedonistic act utilitarianism of the classical, highly reductionist variety expounded by Bentham –
compares highly favorably to the available alternatives. Therefore, my examination of our various moral intuitions will be guided by the assumption that we can probably learn something methodologically useful if we examine the various ways in which the intuitions accepted by utilitarians differ from those they commonly reject.

What are the upshots of these working hypotheses? First, my investigation of our various moral intuitions will focus largely on their functional properties from a methodological point of view. That is to say, when we ask “ought we trust this intuition or not?,” I interpret this question largely as “if we accept this intuition as valid, non-inferential warrant for a moral premise based upon it, would such a premise make our moral theory better able to fulfil the functions we would like it to fulfil?” So that is what I mean when I describe my criteria as functionalist-methodological. Obviously, it is necessary to clarify what I mean by the functions we would like our moral theories to fulfil.

* First of all, there is no denying that descriptive functions are included among those we would like our moral theories to serve. Although I will argue that the importance of descriptive fidelity (1) is being overestimated when evaluating normative theories, it doesn’t follow from this that I think it irrelevant. It is very hard to see how any moral theory could make sense (let alone be accepted as a relevant guideline by anyone) if it didn’t seem to capture a great deal of what was already implicit in our initial preconceptions, so it’s proper to specify at once that my stance is not one of refusal to recognize descriptive success as relevant to the evaluation of a normative theory’s overall merits.

* If we were to think of the functions of a moral theory as distributed along a spectrum from “most descriptive” to “most normative,” we may view the previous function as the one farthest to the descriptive end – simply accounting for how we do judge about morality. On that analogy, we may perhaps view the next step on the spectrum as the function of being morally truth-tracking, seeing as truth – although it is a highly normativized concept – is something that we ascribe to descriptively accurate conveyances of what is the case. And although the term “moral truth” is not uncontroversial, it is hardly controversial to say that if it makes sense to speak of such a thing as moral truth, it then becomes of paramount importance for a moral theory to serve the function of being morally truth-tracking. When I mention this aspect, it is not because I intend to address it much, but rather because it would leave an elephant in the room – a phenomenological lacuna, so to speak – if I passed over it without any mention at all. In this text, the question of what conditions have to obtain for a moral statement to be true or false will
definitely be playing second fiddle to a range of other questions. This omission does not arise from a lack of interest in moral truth, but rather from a lack of certainty about how such a thing could be ascertained. However, the various criteria that I will be concerned with—clarity, universality, consistency, and so on—are criteria that could be uncontroversially presented as likely indicators of an intuition being truth-tracking if any is. So in that sense, it wouldn’t be correct to say that I will not at all concern myself with questions of moral truth.

* Aside from this, there is a second sense in which I will be concerned with moral truth. I will mostly be concerned with pragmatic-methodological criteria, but it wouldn’t be cricket if I refused to tangle with questions of possible, intuitive truth-tracking capabilities at all. Such a refusal would make my overall case too question-begging for comfort. Many central elements of the nonutilitarian position rest on the assumption that it is not unwarranted to assume that many of our moral intuitions can—they don’t have to, but they can—be indicative of objective, moral truths that are valid for other reasons than the methodological considerations I will be emphasizing. Therefore, I will not assert that an intuition’s intrinsic compellingness counts for nothing at all, but rather assume that an intuition’s status as intuitive provides some justification for suggesting that it may be sensitive to moral truth, from which it follows that I will begin with the assumption that every intuition has a somewhat positive presumed baseline credibility (2). This is because—in my view—it is ceteris paribus reasonable to presume in favor of the judgments and notions that do seem plausible. The practical upshot is that I will not hold it warranted to jettison an intuition merely for lacking positive justification along the methodological criteria that I will emphasize. Rather, I will assume that if one is to be warranted in distrusting an intuition, this will have to presuppose either adequate reasons for holding its baseline credibility to be obviated (through a deflationary explanation or the demonstration of irrelevant determinants) or countervailing considerations (i.e., positive methodological unsuitability as opposed to merely a lack of useful qualities, conflict with some other notion that is more plausible, etc.).

* Next, we have the function of promoting precision or clarity (3). We want our moral theories to help us avoid vagueness, to be explicit about the way we interpret “thick concepts,” and to state our positions in as precise and legible terms as possible. And clarity is not just a matter of avoiding vagueness, but also about being explicit. Many moral dilemmas can be elucidated by identifying conflicting premises and ranking them in terms of basicness, many unsound arguments rely on implicit premises that fail to hold up when made explicit, and many controversies arise from a failure to distinguish between conflicting value premises and
conflicting assessments of the facts at hand. Such qualities as these are still *quite far* to the descriptive end of the spectrum (they are certainly relevant descriptive success criteria as well), but we *are* now crossing into the realm of properly normative criteria, seeing as a moral theory will be far more capable of *providing guidance* if it helps us to understand initially-inchoate concepts and assumptions in a clearer way. The practical upshot of this is that if a certain intuition proves well-suited for having its implicit content outlined in clear and precise terms, that will count in its favor. If, in contrast, it proves to be very inchoate and difficult to verbalize, that counts against it.

* We then have the function of *providing a discursive common ground*, or – stated in another way – we see it as a good thing if a moral premise is subject to very wide *consensus*. Oftentimes, moral disputes are exacerbated by a failure to realize that others are operating with conflicting implicit premises, but there are *also* times in which the discussion could have been more productive if someone had successfully identified the premises that all parts were willing to accept. Therefore, we shall hold relative degree of *universality* (4) to be a quality that counts in favor of an intuition’s acceptance, for the less a theory needs to rely on contentious premises, the wider the range of contexts in which it can provide a framework for discourse.

* Next, we shall hold *consistency* (5) to be valuable, seeing as it would be hard to find much use for a regulatory framework based on the deliverances of intuitions that are too capricious and mercurial to seem responsive to any rhyme and reason at all. If, in contrast, we should find that certain considerations correlate very consistently with the verdicts of our moral intuitions, it will not be unjustified to suggest that these judgments may be in conformity with some underlying implicit rule that we may find predictable, reasonable, logically stringent, or otherwise worthy of being taken into consideration as a moral premise.

* We then have *abstract plausibility* (6), which – although it is also a highly relevant criterion from the descriptive viewpoint – often serves a distinctly normative-regulatory role vis-à-vis our initial moral judgments, owing to the fact that people often revise judgments that *seemed initially unproblematic* after finding themselves at a loss to connect said judgments with rationales that seemed plausible on abstract reflection. If somebody says ‘yes, these cases are different in the sense that x and y, but *why should x and y make a difference?’* this typically indicates that they find it implausible upon reflection that the difference in question should matter morally. We shall place a great deal of importance on this criterion: viz., ‘is it possible to demonstrate that the implicit rule of a certain intuition can be understood in light of some
consideration that it seems plausible on abstract reflection to view as morally significant and meaningfully applicable?" In Chapters 3 and 4, we shall be using the term “translation” (in scare quotes, because it’s not the usual kind of translation) quite a bit, by which we mean the process of assessing the abstract plausibility of an intuitive judgment – an intuition that “translates” poorly is thus an intuition that seems to do poorly by this criterion.

* Finally, let us ask ourselves what function could be placed farthest to the normative end of this spectrum? Here is a suggestion: whereas the most-descriptive end of the spectrum encompasses the moral judgments that are being made regardless of whether they ought to or not, the most-normative end of the spectrum encompasses the moral judgments that ought to be made even if (descriptively speaking) they aren’t the least bit in conformity with the judgments that are being made. This last criterion – the ability to make normatively right judgments even when the world is wrong – surely ought to be among the most significant as far as normative ethics is concerned. Now, it ought to be noted that this function is arguably an emergent phenomenon of sorts – seeing as it presumably arises from a theory’s ability to find the appropriate balance between the other criteria (appropriate degree of focus on consistency and abstract plausibility as opposed to descriptive fidelity, and so on.).

This will have to do. Obviously, the list above is not meant to be exhaustive, there undoubtedly being many other relevant criteria beside these. Nonetheless, it provides us with a starting point. For the sake of convenience, let us briefly summarize our criteria, which were seven in number and presented in due sequence as: descriptive fidelity (1), presumed baseline credibility (defeasible by deflation) (2), precision/clarity (3), universality (4), consistency (5), abstract plausibility (6) and then there is that final and seventh, emergent quality that may be a little hard to encapsulate with a very straightforward term. What should we call that one? In the interest of at least having a designation, we’ll simply call it the Seventh Criterion (7).
Chapter 2: Taking sides

- The chapter where I motivate my position, muse on the connection between abstraction and moral progress, and sketch out the initial contours of the stance I will defend.

2.0.0 – Chapter introduction

When presenting my two working hypotheses at the end of the previous chapter, I didn’t yet produce much of a motivation for presuming in favor of the Benthamite view, although I did indicate it being connected with his method’s past successes in satisfying the Seventh Criterion. Put shortly, the primary grounds of my commitments are straightforwardly empirical. If we want to make an assessment of how well a certain methodology or set of premises can help us to make normatively good judgments, we have no better yardsticks available than an assessment of how well it has performed in the past. And as far as my own assessment of the relevant source material is concerned, it does seem as if the utilitarian theory’s track record as a prominent force for social reform counts strongly in its favor.

In this chapter, we shall be looking at a handful of cases. When “cases” are being referred to in the context of moral-intuitional epistemology, the term usually denotes cases that are imaginary. Such cases tend to involve many remarkable elements, like utility monsters, repugnant conclusions, ruthless transplant surgeons, experience machines, unholy constellations of lynch mobs, hapless magistrates and potential scapegoats, and – perhaps the most prevalent element of them all – out-of-control trolleys and the tough choices they impose on bystanders. As diverse as these cases may be, the majority of them still display one unifying characteristic, namely that of being structured in such a way that if they are evaluated in accordance with the methods and premises of the utilitarian theory, the resulting implications will be found unappealing in the extreme. Cases like these are such a significant part of the methodological landscape that I shall be compelled to address them eventually – mainly in Chapter 4, to be specific – but as far as concerns this chapter, it is a very different type of cases we’ll be looking at.

The ability to avoid unacceptable implications in imaginary cases is very much a descriptive type of success criterion. To call these implications unacceptable is a descriptively correct statement about the pertinent psychological facts. Although these descriptive facts are significant and relevant in many ways, they cannot by themselves provide a basis for determining if the implications in question are normatively unacceptable. As already indicated, I will not
deny that a theory’s failure to guarantee against unacceptable implications is relevant, but I fail to see why it should be ascribed such a high degree of relevance as seems to be the prevalent assumption. This defect is one that counts for quite a lot when evaluating a theory’s merits as a descriptive account, but the primary emphasis of most ethicists falls within the normative side of the spectrum by a comfortable margin. And as long as we are concerned with a theory’s normative merits, it is hard not to think that there are several other criteria – we outlined quite a few in 1.2.4 – that should count for considerably more than descriptive fidelity. In this section, we shall begin by looking into that seventh and last success criterion I presented: the ability to arrive at normatively good judgments even when they are at odds with what is – descriptively speaking – the correct view of the day. Is it a theoretically desirable quality to accommodate the notions that we do – as a matter of fact – hold to be correct? Yes it is. But is that quality anywhere near as desirable as the ability to unmask and contradict those notions that – although they accurately convey the pertinent facts about the moral consensus of the day – are nonetheless and unbeknownst to the majority normatively incorrect by some other, less transitory standard that in due time and the eyes of posterity shall be recognized as more worthy of consideration? One must be careful when using the word “obviously” in philosophy, but as far as this particular question is concerned, I find it hard to imagine any grounds for denying that the answer is not merely negative but obviously so.

That is why I find myself so perplexed by the great interest directed at imaginary cases, and – more so – the comparative lack of emphasis on cases that are historical. These are often disputes where we now have very clear, confident and intersubjectively consistent views about which position turned out to be the better one in the long run, whereas the people who were party to them at the time had no way of knowing which horse should eventually emerge as the wiser bet. This makes historical cases empirically informative in a way imaginary cases are not, seeing as the people back then were not in a position to tailor the application of their premises so as to arrive at the implications they knew beforehand to be the desirable ones. This makes such disputes highly promising contexts for evaluating the various arguments, premises and methodologies involved, seeing as the verdicts made on their basis are far more likely to provide an accurate indication of how well these theoretical resources perform when – if one may say so – having to “operate in the blind” and under ecologically valid conditions. But enough of this. Let’s have a look at these cases, and see if we can learn something interesting.
Section 2.1 – Hindsight is 20/20: A few object lessons from history

2.1.0 – The presumed incongruities of Mr. Mill’s proposal

For our first case study, we wind back the timeline to 1859, at which point John Stuart Mill has just gone forth and put some real crankery into the public sphere. His newly released Thoughts on Parliamentary Reform marks the first time he goes public with his advocacy of women’s suffrage, and – with the past being the foreign country it is – the proposal is found as preposterous by Mill’s contemporaries as it seems long-overdue to the modern reader. There are some interesting features to note about the rebuttals. The naysayers do not seem to field much in the way of formal arguments, but neither do they seem very embarrassed by their absence. Although there are some objections on practical grounds, such concerns play second fiddle to flat-out rejection on account of “the unfitness and impropriety of allowing women an active share in public affairs” (quoted by Nicholson 1998, p.472). The Times finds it a “complete inversion of the natural order” to “base the franchise upon principles half recognised by a few philosophers, but utterly strange to the mass even of educated men” (p.471). In some cases, it is graciously acknowledged that there probably are valid and reasonable arguments in favor of his position – the Saturday Review concedes that “when Mr. Mill makes a legislative proposal, something may probably be said in its defence” (ibid.). But this doesn’t seem to matter much. The fact that the proposal is found so clearly and immediately ridiculous is held up as sufficient reason in itself for rejecting it. Returning to the Saturday Review, it is remarked that “If an arrangement strikes ninety-nine out of a hundred men as supremely ludicrous, there is probably some real incongruity in the plan itself” (ibid.). Stated in another way, it’s being admitted that although there may be good arguments in favor of the proposal, these are undoubtedly outweighed by the good counterarguments that – while not explicitly presented – must surely exist on account of how preposterous the proposal is found by the majority. It is not denied that Mill’s conclusions follow from premises that are hard to fault, but rather than taking this as a reason to feel less rock solid in one’s convictions, it is taken as cause for lamenting the “kind of pedantry which… occasionally disfigures Mr. Mill’s writings” (p.471).

Since our views have changed so dramatically on this matter, it is now easy to see how silly it was to take one’s own immediate, intuitive rejection as a surefire symptom that valid

---

6 It was, for instance, objected that the whole measure was pointless since enfranchised women would just end up voting the same as their husbands anyway. So there’s that.

7 With Mill being an esteemed figure in a polite age, there is little of the outright japery one instinctively expects whenever someone proposes to move the world forward.
counterarguments surely must exist. But from a structural point of view, this “no-smoke-without-fire” maneuver is not dissimilar from the still-widespread practice of holding that strong intuitive judgments should be given the benefit of the doubt even when they prove notoriously hard to justify with convincing moral principles. Come to think of it, I even indicated (in 1.2.4) that I myself regarded it as valid, so it’s not as if I am in a position to pontificate.

2.1.1 – The intolerably obvious realities of Professor Rashdall’s day

Even within the utilitarian camp, many have found it hard to accept the stark and reductionist axiology that is Benthamite hedonism. Hastings Rashdall – one of the earliest proponents of the more salonfähig, anti-Benthamite stance that would in time be called ideal utilitarianism – is writing at a far more recent point in history than the newspapermen who rejected Mill’s call for universal suffrage, but the succeeding century has already contained more than enough changes to make his convictions take on an archaic air. At first, the anachronisms seem to be of the rather innocent sort, as when he regards it as quite definite disproof against Benthamite hedonism that such an axiology could not properly account for the undeniable and unconditional value of sexual purity and the institution of marriage, or for the fact that one intuits binge drinking to be inherently less permissible than patterns of alcohol use held more injurious by the prevailing medical wisdom of the day\(^8\) (Rashdall 1907, pp.197-203). So far, it doesn’t seem as if insisting on a fundamental distinction between higher and lower pleasures leads to anything more terrible than a little priggishness. But it takes on a more sinister aspect as the book progresses. Rashdall takes a nuanced view of how much inequality can be tolerated for the sake of a minority’s opportunity to pursue rarefied, higher-pleasure pursuits like the study of classical languages, and admits that in these cases it’s not entirely self-evident that the Benthamite view ought to be jettisoned after all (pp.234-238). However, he continues by noting that there are other cases that demonstrate the untenability of Benthamite hedonism in far more indisputable terms. Noting that the case he now presents is one “in which probably no one will hesitate” (p.238), he continues by noting that:

It is becoming intolerably obvious at the present day that all improvement in the social condition of the higher races of mankind postulates the exclusion of competition with the lower races. That means that, sooner or later, the lower Well-being – it may be ultimately the very existence – of countless Chinamen or negroes must be sacrificed that a higher life may be possible for a much smaller

\(^8\) Ironically, that particular example is one in which clinical consensus has since come to align with Rashdall’s intuition. Nowadays, “two glasses of port every day after dinner” (p.202) are no longer considered more detrimental to health than a single drinking binge per month, so in this case it seems there really was no smoke without fire.
number of white men. It is impossible to defend the morality of such a policy upon the principle of equal consideration taken by itself and in the most obvious sense of the words. If we do defend it, we distinctly adopt the principle that higher life is intrinsically, in and for itself, more valuable than lower life, though it may only be attainable by fewer persons, and may not contribute to the greater good of those who do not share it (Rashdall, 1907, pp.238-239).

The modern reader will surely agree with Rashdall that such a policy is “impossible to defend” on the basis of a strictly Benthamite axiology, but in contrast with the well-bred Oxford don, our contemporaries are more likely to view such an incompatibility as a notable point in the theory’s favor than a fatal defect. Yet from the viewpoint of his own society, Rashdall is making a straightforward appeal to counterintuitive implications in order to argue against a theoretical view. He is pointing out that if Benthamite hedonism is correct, this would lead to the highly counterintuitive implication that it is not morally justified for Europeans to expand and flourish at the expense of the darker-hued multitudes. He does not give any technical, premise-to-conclusion account of why such an implication must be rejected. He simply states – as was quite correct at the time – that this happens to be the view of the overwhelming majority. As in the previous example, we see once again that although the substantive opinion being championed is now viewed as beyond the pale, the *structure of the argument* is fully alive and well. One identifies a certain implication that would follow from a set of premises, points out that the implication in question is found unacceptable by nearly all, and concludes that this demonstrates the untenability of the premises. Whereas one may describe the tendencies of the previous example as an attempt to treat one’s intuitions as valid signifiers of the presence of as-yet-unknown rationales that are valid, Rashdall’s argument can be seen as more of a straightforward invocation of what we might call “epistemic force majeure” – that is to say, asserting that some certainties just happen to be ineradicable facts on the ground about how we do think and must think, and that a theorist failing to accommodate them is simply banging his head into the wall by expecting the terrain to conform to the map.

2.1.2 – Lord Devlin’s moving and welcome humility

We make another jump forward as we move our focus to the latter half of the 1950s. It has been nearly a full century since Mill’s contemporaries despaired over his doctrinaire and idiosyncratic views, but his indirect influence remains alive and well within Britain’s life of the mind. Such influence is readily apparent as Sir Patrick Devlin – eminent jurist and later to become Lord Devlin – peruses the recently published report of the Wolfenden Commission. The Commission – tasked with giving its recommendations for reform of the nation’s vice laws – has raised many eyebrows with its suggestion, very much out of place in the fifties, that
“homosexual behavior in private should no longer be a criminal offence” (quoted by Devlin 1972, p.226). Whereas Rashdall and the Victorian newspapermen were caught, as it were, in flagrante delicto as they expressed now-unfashionable views on current issues of their day, Devlin avoids this fate as far as the substantive question is concerned. When he takes issue with the Report, it is not so much on account of the recommendations it makes, as the theoretical presumptions that underlie them. More precisely, he takes issue with the Commission’s assertion that mere public condemnation in itself can never suffice to justify criminal prosecution.

The Report’s authors state their grounding principles in terms bearing the clear stamp of Bentham and Mill’s influence, as they explicitly assert that the censoring of immorality as such has no place among the functions to be served by the criminal law. Although a great many acts deemed criminal tend to overlap with acts deemed immoral, the question of immorality on its own is neither here nor there as far as the justification of punitive intervention is concerned. After listing the various types of harm that the criminal law rightly ought to protect the public from, they continue by making it clear that “there must remain a realm of private morality and immorality which is, in brief and crude terms, not the law’s business” (ibid, p.226). Aside from what is “necessary to carry out the purposes we have outlined” – that is to say, upholding public order and preventing exploitation and the like – the Commission holds that it is simply not within the criminal law’s legitimate sphere of application to “enforce any particular patterns of behavior” (ibid). This statement of intent is one that most people will find easy to endorse at first – Devlin refers to it as “an admirable and modern starting point” – yet, as often happens with premises that seem unobjectionable in the abstract, their implications can take on a troubling air when outlined more clearly. The Commission didn’t just state that many forms of immorality are outside the law’s proper sphere of application, but rather that no forms of immorality at all merit legal intervention qua being immoral. The corollary being that unless one can justify prosecution of an activity on the basis of preventing harm, the fact that we consider it immoral – no matter how strongly or deeply or universally – counts for nothing at all. Certainly, it’s not hard for a modern reader to endorse this strict separation of crime and immorality as long as the particular case involves putting people in jail for being gay. But the premise isn’t just about that particular case: it is about any and all cases where society lacks any harm-based justification for intervening in activities that are found morally repugnant in the extreme. At this point, it may appear a little strong. Nothing at all? Entirely beside the point? We do live in a democratic society, so it’s not immediately obvious that we have no legitimate reasons for expecting our institutions to display an ever so slight sensitivity to the core values that bind us together.
It was on this basis that Devlin – initially approving of the Report – gradually came to view its premises as deeply problematic (Dworkin 1966, pp.987-998). Expounding on them in his 1958 Maccabean Lecture, Devlin takes issue with how the Commission seems to presuppose that the criminal law’s sole function is to protect the citizenry from a range of harm-inflicting behaviors that just happen to coextend with behaviors that are morally condemned by said citizenry, whereas the condemnation itself is in a sense epiphenomenal and means nothing when directed at something the law sees no need to protect anyone from. And there is indeed something unpleasantly technocratic and paternalistic about this notion – that a legal code’s sole concern is to prevent behaviors that have been judged harmful to the citizenry, without in any way being responsive to the very same citizenry’s steadfast convictions about what behaviors they are and are not willing to tolerate in their midst. The flip side of this democratic argument, however, is the possibility of mob rule and arbitrary persecution that results if one does hold it legitimate to criminalize anything on the sole ground of public antipathy. Devlin is aware of this possibility, as he is careful to specify that not just any old antipathy will do:

It is not nearly enough to say that a majority dislike a practice; there must be a real feeling of reprobation. Those who are dissatisfied with the present law on homosexuality often say that the opponents of reform are swayed simply by disgust. If that were so it would be wrong, but I do not think one can ignore disgust if it is deeply felt and not manufactured. […] matters of this sort are not determined by rational argument. Every moral judgement, unless it claims a divine source, is simply a feeling that no right-minded man could behave in any other way without admitting that he was doing wrong. It is the power of common sense and not the power of reason that is behind the judgements of society. […] We should ask ourselves in the first instance, whether looking at it calmly and dispassionately, we regard it as a vice so abominable that its mere presence is an offence. If that is the genuine feeling of the society in which we live, I do not see how society can be denied the right to eradicate it (Devlin 1972, pp.233-234).

So it’s not a complete free-for-all in the mob rule department: one cannot persecute people on the basis of disgust unless it’s “deeply felt and not manufactured,” and the decision to do so has to be made in a calm and dispassionate state. All good, but the upshot remains that Devlin ends up positing it as a democratic right to enact laws against people simply because the majority cannot stand them, with no further reason strictly necessary.

This seems to have been a guiding refrain throughout all three of our cases: the notion that sometimes – when a moral position is sufficiently popular – challenges against it can be discarded without owing anyone a reason for doing so. This notion may be congenial to some – The Times responded to Devlin’s lecture by praising the “moving and welcome humility in the
conception that society should not be asked to give its reason for refusing to tolerate what in its heart it feels intolerable” (quoted by Hart 1972, p.242) – but to an ethicist, it ought to strike a fundamentally sour note.

2.1.3 – Where are we headed with this?

Now, this is the kind of cases that I think moral philosophers should be concerned with. Notions like moral truth and moral progress are sometimes found a little contentious in the abstract, but it would be challenging to find someone unwilling to concede that it does make sense to regard the differing stances in these disputes as more and less normatively successful. To get it right even when the world gets it wrong: that is the core of what I tried to capture with the admittedly-inchoate function I had to settle for calling the Seventh Criterion, which – although it is a spectacularly uninformative designation – does have a certain mystique going for it. But we must proceed. These cases were not so much presented for their amusement value as for the useful things we hoped to learn from them. We must be more specific.

Whatever else may be said for the classical utilitarian approach, the cases did seem to indicate that there is something it keeps getting very right. What could that be? We must look at the commonalities between the cases. First: in all three cases, the now-disfavored stances were defended by appeals to intuition – more specifically intuitions of the type we classified as (b.) (notions that were general but didn’t show much promise as formal premises) – and they all seemed to presume the evidentiary adequacy of such intuitiveness. We’ve addressed this a little bit already, classifying the various justificatory presumptions as “no-smoke-without-fire,” “epistemic force majeure” and – actually, we didn’t come up with a name for Devlin’s approach, but we can dub it “moving and welcome humility.”

Can we derive something interesting from this? Perhaps not so much, actually. As far as contemporary ethicists are concerned, even the most intuition-friendly among them are unlikely to ascribe evidentiary adequacy to intuitions, and as for the mostly-implicit rationales for trusting them, it may be rather hard to argue against according them some relevance. They did seem to mirror rationales that I myself accepted as valid in 1.2.4, such as presumed baseline credibility, universality and descriptive fidelity. The cases may indicate that these justificatory presumptions are problematic, but they didn’t seem to provide a sufficient basis for rejecting them. Maybe we can find some other commonality that gets us to a more interesting place?

Here is a commonality that may be promising: all of these cases were concerned with distinctions. The defenders of the status quo all insisted that certain things were relevantly
different without presenting any explicit rationale for that position. They seemed to insist on ascribing some sort of normative significance to the difference between male and female disenfranchisement, the higher pleasures of Europeans and the lower pleasures of “Chinamen and negroes,” and “mere emotional” as opposed to “deeply felt” disgust. It is a very central complaint against utilitarians that they fail to account for many relevant differences, and that they are too reductionist and given to overgeneralization. But if we are to judge on the basis of these cases, it seemed as if those tendencies were pretty conducive to good normative judgment, and that the obverse tendencies – what is often called nuance or a salutary avoidance of oversimplification when it doesn’t run afoul of history like here – seemed similarly obstructive.

Here is a phrase we briefly alluded to before (in connection with the abstract plausibility criterion): “why should that make a difference?” This unassuming phrase may well be viewed as one of the primary drivers of moral progress, seeing as the discontinuation of unjust practices is so often spurred by the realization that they weren’t after all relevantly different from other practices already recognized as unjust. To point out relevant similarities between practices held respectively acceptable and unacceptable is one of the most widespread strategies of moral reform, and – conversely – it also seems as if an insistence that the practices in question are not comparable plays a similarly vital role in obstructing the former. At this point, we seem to have found a promising element. We’ll proceed to the next section now, we’re we’ll be looking at how this consideration connects with another concept that will play a very major role in this text.

Section 2.2 – Reflections on abstraction

2.2.0 – Apples and oranges

We came to focus on the matter of distinctions as promising. In this subsection, we’ll be looking at the way this element interacts with another concept that will concern us greatly throughout this text, namely abstraction.\(^9\) The utilitarian approach seemed strongly geared

---

\(^9\) What do we mean by abstraction, anyway? Although the terms abstract and concrete are understood qualitatively in some contexts, it functions better for our purposes to regard the concrete/abstract pair as describing a difference in degree rather than in kind. Something is more concrete if it is more closely tied to a particular context and more abstract if it is more divorced from any particular context. A particular act of theft is very concrete. The general action class of theft is somewhat more abstract, but less so than an account of how the action class of theft relates to the Categorical Imperative. And this account in turn is still more concrete than the grounds of the principle itself. So it’s a sliding scale rather than two clearly delineated categories. Our intuitions about dictators and abortions are concrete compared to our intuitions about oppression and personhood, but abstract compared to our intuitions about Pinochet and Roe v. Wade. Yet there is mutual influence in all directions. Our intuitions about abortion may partially be influenced by our abstract intuitions about personhood.
towards favoring the more abstract forms of intuitiveness (c. and d.). But there is also a
different, perhaps subtler sense in which the theory seems to favor that which is abstractly
intuitive.

Many philosophical conundrums arise from the way different and often conflicting
notions seem intuitive from an *abstract* perspective as opposed to one that is more *concrete*.
Metaphysics is a field where we find particularly clear examples of this tendency. The
concrete intuition is to accept diffusely bounded terms like “heap” and “crowd” as meaningful
and integral to our conceptual schemas, because our everyday point of view accepts vague
boundaries. But in the abstract, this vagueness doesn’t sit comfortably. From that perspective,
we can no longer hand-wave away the conviction that the boundary between heap and not-
heap must correspond with *one particular number*, which seems intolerably arbitrary once we
try to decide precisely which one should mark the transition. In this case, many will favor the
abstract intuition, and accept that soritical predicates may just be terms of convenience that
are not meaningful in the deeper sense of carving Nature at the joints. Other times,
philosophers favor the concrete intuition. In the abstract, it seems unacceptable to designate
some arbitrary point at which a quantity can no longer be subdivided. But when the concrete
implications of infinite divisibility are outlined in Zeno’s Paradox, they do not undermine our
intuitive certainty that this is not how time and distance interrelate in the real world. The
paradox is acknowledged as puzzling and hard to account for, but no one suspects that we are
actually mistaken in our concrete intuitions about how the race between Achilles and the
tortoise would play out. And so it goes. Determinism can often seem like a plausible position
on abstract reflection, whereas the notion of free will is impossible to dispel when we take a
first-person perspective and ask ourselves which implicit presumptions *must by necessity*
underlie our engagement with the world. The man in the street is a Cartesian, while the
theorist in his armchair is liable to think substance dualism raises more questions than it
answers.

There is a *pattern* to be seen here: it often seems most intuitive to operate with
fundamentally disjunct categories when taking a first-person, context-embedded type of view,
whereas such sharply delineated boundaries tend to seem counterintuitive when examined
from a depersonalized and decontextualized perspective. From the concrete perspective, it
seems unobjectionable to view the various phenomena of the world as fundamentally separate

---

and partially by our concrete intuitions about *Roe v. Wade*, while themselves having an influence in turn on our
intuitions both at the higher and lower level of abstraction.

38
and distinct from each other – mind/body, reasons/causes, heap/not-heap, sublunary/superlunary, air/fire/water/earth, brutes/men, good/evil, earth/turtles-all-the-way-down, and so on. Not so much from the abstract perspective. When we switch our thinking into that mode, it often seems more right to interpret the world’s seeming plurality as surface phenomena of a deeper, interconnected substratum, and the mind inclines towards monism and atomism. It no longer seems so intuitive to think that all sorts of everyday categories should be fundamentally separate and different. They do interact, after all, and a lot of the familiar boundaries now seem to take on an air of arbitrariness. Sure – water and fire seem pretty different, but does that have to mean they can’t just be different configurations of the same building blocks? As long as there’s causal interaction between the mind and the body, they can’t be that fundamentally different, right? And if there’s Eternity before Creation and Eternity after the End of Days, why not cut the middleman by presuming Eternity all the way with no need for a bounded Present squeezed between two slices of Eternity? The tendency on abstract reflection, it seems, is to do away with absolute boundaries and fundamental differences in kind.

We see these conflicting tendencies reflected in the way abstract theories always try to account for phenomena with as few and wide-reaching principles/concepts/laws/forces as possible, whereas our everyday phenomenology is primed to operate under the implicit presumption that the things around us are irreducibly plural. Our concrete intuitions prefer to split things up into manageable chunks, while our abstract intuitions tend to see everything as continuous and interconnected beneath the surface. The former are primed to detect salient differences, whereas the latter are sensitive to underlying similarities.

Our concrete intuitions are unhappy with notions that diverge too much from the surface phenomena of the world as experienced by us, and for good reason: that world-as-it-seems is precisely the world they have been calibrated for, and their purpose is one of practical guidance within familiar parameters. Not so with the intuitions elicited by contemplation of abstract, speculative matters. These judgments arise from different processes with different functions, and are not burdened by the same strictures as their more earthbound counterparts. Freed from this need to uphold resemblance with our everyday phenomenology, our abstract intuitions find themselves responsive to demands of a different sort, like the desire for parsimony, symmetry and consistency, a reluctance to accept vagueness, a distaste for the arbitrary, and – related to the latter two – a dissatisfaction with rationales that evade precise verbalization. Thus, we keep finding that notions that appear foreign and absurd from
a grounded, “first-person” perspective end up seeming compelling and logical when assessed from, as the Sidgwickian phrase puts it, “the point of view of the Universe.”

Stated more colloquially, it seems that our intuitions about concrete matters favor a “don’t compare apples and oranges” perspective, whereas our intuitions about abstract matters are more likely to indicate that “fundamentally speaking, they’re both fruit, aren’t they?” Our concrete intuitions are like the person who says “sometimes, you just have to accept that things are the way they are with no deeper explanation,” whereas their abstract counterparts are like the person who won’t listen on that ear. When our abstract intuitions complain about some inconsistency or other, our concrete intuitions will insist that “You can’t compare this case with that! They’re not equivalent at all! Don’t compare apples and oranges!”

Or maybe we should use a more concrete illustration, such as “You can’t compare male disenfranchisement with female/lour higher pleasures with their lower pleasures/mere emotional disgust with profound disgust. Don’t come here and mix apples and oranges with that implausibly reductionist theory of yours!” Ah. That’s right. Now we are back on track with the distinctions again. Here’s an elaboration on the second of my working hypotheses from 1.2.4: when the utilitarian method displays such a remarkable tendency to be proved right by history after initially stepping on the toes of commonsense morality, this is because it constitutes an unusually successful attempt at devising the kind of premises and methodological ground rules that can prevent us from sliding into an “apples-and-oranges” mode of intuiting whenever we are tempted to evade following our “all fruit” intuitions all the way to a substantive conclusion that – although in actuality a very sound position – is found too unpleasantly unorthodox.10 Now, this regulatory function – a theory’s ability to keep us intuiting in an “all fruit” mode even when concerns for descriptive fidelity strongly tempt us in the direction of “apples and oranges” – is a function that Bentham’s theory has, in my view, yet to be bested at performing.

---

10 This is why I always get skeptical when people advocate taking a “toolbox” attitude towards the various normative theories, by using some utilitarian reasoning and some deontological reasoning and some virtue ethical reasoning all in accordance with what the particular issue “requires.” This kind of approach tends to sound very reasonable and moderate – as opposed to a doctrinaire insistence on using one set of tightly linked premises for every moral issue – but it’s very hard to see how one could guard against letting it slip into a cherry-picking mode where additional, supplemental premises from other parts of the “toolbox” are invoked whenever one is tempted to evade the implications of the current set.
Section 2.3 – Finishing touches before we proceed

We are now ready to outline our initial position on the minimalism-vs-maximalism dichotomy, and our strategy for grounding that position.

2.3.0 – Putting the finishing touches on our initial case for intuition-minimalism

Let’s look at a definition of “intuition” that was not involved when we did our conceptual analysis in section 1.1 – because it was too broad and strictly functional to be useful back then – but that is now worth having a look at. The definition, coming from Gutting (2016, p.319) is simply “premises that require no argument.” Elegant and to the point as far as methodological function is concerned, although of course encompassing a far broader category than what we mean by intuition as a class of mental event. When we involve it here, it is because it touches very directly onto a point that is now important: whenever we do accept a premise qua being intuitive (or reject it for leading somewhere counterintuitive), we also narrow the range of what is subject to argument. If we base a premise on intuitiveness, then – well, people can accept it or they can refuse to accept it. But there’s not much of a role for arguments to play. Back in the historical cases, the defenses of the status quo positions we now consider misguided all seemed to involve variations on the claim that further reasons are not strictly needed for rejecting implications that seem too counterintuitive. So, the more we permit the use of intuitive warrant (or diswarrant, as may be), the more we extend the parts of the discursive space in which it is permissible to abstain from presenting reasons for one’s position. We certainly saw from the historical cases how pernicious this kind of license can be, and – conversely – it is hard to see how it could ever be a good thing. And since intuitions are characterized by the introspective inaccessibility of their grounds of justification (provided there are any to be found at all), it’s generally very hard to discover when they are misleading. So there are both discursive and epistemic considerations that count against intuitions in a very basic way. Because of this, we adopt the position that – other things being equal – it is always better to permit less intuitive warrant rather than more. Less is more! So we’ll be arguing in favor of doing it in the axiomatic way, the one we described as more “physics-style” and privileging the more-abstract types of intuition (c. and d.). That’s because we won’t have to rely on as many intuitions if the ones we do use are highly abstract and general, and also because this approach – especially in its utilitarian manifestations – appears to display a remarkably good track record as far as concerns the normative-methodological desideratum we dubbed the Seventh Criterion.
So our position is intuition-minimalist. We must therefore outline the first rudiments of our strategy for facing the Only Game in Town objection, as outlined in 1.2.3. That will be an ongoing task throughout the entire text. We begin by looking at the initial leads we have collected throughout this chapter.

**Lead #1: Abstraction.** It seemed as if the approaches that functioned well in the historical cases were characterized by privileging the kind of more-abstract intuitiveness that favored the “all fruit” perspective over “apples and oranges.”

**Lead #2: Back to Bentham.** Closely connected with this, the utilitarian theory seemed to have a lot going for it – more than many would assume, perhaps – and it was also very in sync with the kind of abstraction-favoring approach we have sided with. And since the utilitarian theory is highly reductionist, it seems like a good starting point for pursuing our goal of minimizing the use of noninferential premises that would have to rely on intuitive warrant.

### 2.3.1 – The obstacle we have to address next

In light of the considerations presented in this chapter, one might have thought things were going quite swimmingly already. The less-abstract intuitions already seem to have quite a bit of mud on them, don’t they? Might not this be enough already to assign them less evidentiary weight than their abstract rivals, which display little in the way of such defects?

That would have been premature. At this point, there is one particularly significant obstacle in the way. Since we are building our case on the interconnected threads of “favoring abstraction” and “favoring the utilitarian approach,” we have to be sure that we really can justify the claim that such an interconnection can be meaningfully referred to. Do utilitarians really rely on intuitions that are more abstract and less concrete than the adherents of less-minimalist, rivaling approaches? This question is problematic, because it’s *not really that clear* when an intuition counts as one or the other. Granted, it’s not hard to argue that intuitions about *particular cases* are concrete. But those aren’t *really* the kind of intuitions that deontologists and other non-utilitarian ethicists place their primary reliance on. They will often examine and compare case intuitions in order to look for underlying patterns that may be informative on a deeper and more general level, but it is *that* deeper level they are interested in. They most certainly do not go around invoking absurd premises like “The Principle of Trolley Case 1” or “The Doctrine of the Unacceptability of What That Surgeon Did” when they engage in moral theorizing. What they *might* do is refer to those cases in order to point out that they display
certain interesting features that may be informatively interpreted in light of some principle that in itself is eminently abstract. Old deontological standbys like the Doctrine of Double Effect and the Doctrine of Doing and Allowing are perfectly abstract in their own right, and many find them intuitively compelling as general notions.

On top of that, it’s not as if the utilitarian theory is never on thin ice from an abstractly-intuitive point of view. There are times when it takes some pretty extensive liberties with our very basic moral notions, and the concept of utility is one that runs into problems at the abstract level if one tries to quantify it, seeing as this leads to that abstractly-counterintuitive kind of gridlock where things either seem too vague or else they seem too arbitrary. And on the flip side, the system also derives a great deal of its plausibility from its success in accounting for a very wide range of case judgments and general notions that seem implicit to our commonsense morality. So if we are going to pursue that kind of “abstract=utilitarian/concrete=deontological” line of thinking – for it does seem to have some initial plausibility – it would first be necessary to clarify and develop our notions about the way abstraction and intuitiveness relates to each other.

So there is a problem. The next step is to proceed to Chapter 3, where we shall address it head on with an attempted expansion of the intuitional-epistemological terminology.
Chapter 3: Distinctions

- The chapter where I examine intuitions, introduce new terminology, and attempt to distinguish between more and less abstract intuitions

3.0.0 – Chapter Introduction

Towards the end of Chapter 2, we seemed to be forming a position of sorts. We had concluded that there were good reasons to keep the reliance on intuitions minimal (2.3), and – based on an admittedly eclectic mixture of historical (2.1) and phenomenological-speculative (2.2.0) considerations – set out to cobble together an idea of how to justify such a minimalist approach in the face of what we called the Only Game in Town objection.

But then we hit a snag. Our gradually forming argument had become intertwined with a postulated connection between utilitarianism and abstract intuitiveness, but this seemed in need of clearer and more definite demonstration. Even if we did have a hunch that the intuitions relied on by utilitarians could in a sense be called “more abstract” than the ones they often dismiss, it wasn’t clear how meaningfully the abstract/concrete dimension could at all be applied to the intuitions that do serious work in contemporary ethics. If we were going to pursue this line of argument, it would be necessary to form a clearer idea of how – or whether – it could make sense to describe intuitions about general notions as comparatively more and less abstract.

And that is where we are now. Towards the end of this chapter, we’ll be outlining a proposed taxonomy that distinguishes not just between intuitions that are concrete (about particular cases) and abstract (not about particular cases), but also between those intuitions that are abstract to a high degree and those that are abstract to a lesser degree. But in order to do that, we must first have a look at what it could mean when we talk of the intuitions that utilitarians rely on and dismiss respectively. An examination of these questions will be the subject of sections 3.1 and 3.2, whereafter section 3.3 outlines the new taxonomy.

In Chapter 1, we distinguished between four classes of intuitions: intuitions about cases (a.), general notions that are found intuitive (b.), methodological intuitions (c.), and general notions that are found intuitive both in the substantive and methodological sense (d.). In what category shall we place the intuitions in this chapter? Oftentimes several and overlapping ones, seeing as the implicit rule observable within a pattern of normative judgments can also be regarded as intuitive when made explicit and evaluated as a general notion. Thus, negative hedonism can both be considered an intuition in the sense that it seems implicit in many case
judgments (a.) and an intuition in the sense that when the content entailed by these case judgments is outlined in declarative terms, most will regard it as an intuitively-plausible general notion (d.).

It is worth noting that this chapter will by and large be concerned with a descriptive-classificatory comparison of these intuition classes, whereas the evaluative comparison – that is to say, an evaluation of their relative epistemic and methodological merits in accordance with the criteria from 1.2.4 – will be the subject of Chapter 4. This is not to say that the criteria won’t be as much as mentioned, but – in the interest of avoiding overlap – this chapter will mainly be concerned with the characteristics that could make it meaningful to describe one class as more or less abstract than the other.

Section 3.1 – The intuitions that utilitarians rely on

3.1.0 – What do we mean by “utilitarian” intuitions anyway?

Would it be correct to describe the Principle of Utility itself as an intuition? That is to say, is the Principle that “approves or disapproves of every action whatsoever, according to the tendency which it appears to have to augment or diminish the happiness of the party whose interest is in question” (Bentham 2007, p.2) an example of the sort of thing we’ll be referring to when we refer to intuitions in this chapter?

Not really. Although many will agree that the Principle can be called intuitive, this does not entail that it is an intuition – or at least not in the sense most interesting to us. It seems more right to say that the Principle is found intuitive because several of its component propositions are intuitive. As for the general notions I view as supplying the grounding premises for the utilitarian theory, they are such that they can both be called intuitive in the sense of seeming implicit to many of our case and methodological judgments (a. and c.) and in the sense that when those implicit rules are outlined in declarative terms and evaluated as general notions, these notions are likely to be regarded as intuitively plausible (d. in this case). Later in this chapter, we’ll be looking at intuitions that have a harder time seeming plausible once they are evaluated as general notions, despite seeming to be implicit in patterns of judgment that appear highly intuitive at case level.

What intuitions can be said to underlie the Principle of Utility? First, there are the intuitions that pleasure is a sui generis good and pain a sui generis evil. We’ll dub these positive
hedonism and negative hedonism respectively. Since hedonism is typically referred to as a single axiological premise, it may not seem immediately necessary to classify the two intuitions separately. However, it is unlikely to take much introspection before concluding that negative hedonism is a significantly stronger intuition than its positive counterpart. Other things being equal, we are likely to intuit that someone who opts to bring about \( n \) units of pleasure rather than averting \( n \) units of pain made a morally bad priority. So even if we favor parsimony, it seems proper to say that this premise reflects two separate intuitions rather than a single one.\(^\text{11}\)

It is useful to note that although these intuitions are not experienced as equally strong, this phenomenological asymmetry is not taken into account by the Principle itself. It only states that one ought to maximize overall utility, without making a distinction between the prevention of pain and the promotion of pleasure. This lack of distinction seems counterintuitive when its implications are contemplated at case level, but it seems less counterintuitive when contemplated in the abstract. Is there any positive justification for calling pain and pleasure fundamentally different? Might it not be more likely that we intuit this way as a rule of thumb because pain typically comes in greater quantity than pleasure? If they are radically different, how could they be balanced against each other? And if they are merely weighted differently without being incommensurable, what would the precise exchange rate be? This looks like another of those distinctions that become either too vague or too arbitrary when pondered in the abstract. We recognize this from 2.2.0: differences in kind tend to seem far less intuitive when contemplated in the abstract and away from particular contexts. As one might have expected, Bentham has prioritized abstract intuitiveness above its case-level counterpart when formulating the principle.

Chapter 2 implied that this tendency to prefer incremental holism above disjunct pluralism may itself be seen as the manifestation of an intuition of sorts. Is it a separate, free-standing intuition to prefer holist incrementalism, or can it be subsumed under a more wide-reaching intuition? It seems right to say that it can be subsumed under a more general tendency to favor as much parsimony as possible. We might call this intuition less is more. Can the content of this intuition be stated in propositional form? That it can. Aristotle has already done it for us in the Posterior Analytics, wherein he counsels us that “We may assume the superiority ceteris paribus of the demonstration which derives from fewer postulates or hypotheses” (Baker 2013).

\(^\text{11}\) Furthermore, it can be argued that some philosophical systems – Buddhism and possibly Epicureanism – operate with negative hedonism as their sole axiological premise.
Then there is the element of agent-neutrality. Is this a single intuition or more than one? Is it intuitive *sui generis* that everyone counts for one and no one for more than one, or can this be subsumed under a more general and wide-reaching intuition? In terms of abstract intuitiveness, more generality seems preferable to less. Perhaps one could formulate the content of the underlying intuition as “The onus of justification always falls upon the one who would treat cases differently rather than similarly.” This would cover agent-neutrality as part of a wider scope, and there’s no need for a more specific intuition when a more general will do. We need a name for this intuition as well, so let’s go with *same rules apply*.

What can we say so far? Is it right to say that these four intuitions can properly be described by that term? It does seem so, for it would be hard to imagine an argument that could change the mind of someone who did *not* find them intuitive on their own merits, and yet the overwhelming majority needs no convincing to begin with. It is uncommon to see the initial premises of utilitarianism faulted for being *unconvincing*. The objection is rather that there are also other premises beside them that are convincing enough to merit a place at the same table. No one says that one should *not* treat equal cases equally. Instead, they argue that the utilitarian doctrine fails to account for certain considerations that *make* many supposedly equal cases legitimately different. It’s unlikely to hear someone say that one does *not* have independent moral reason to prevent and refrain from inflicting pain. They may argue that the applicability of the premise is limited by certain retributive considerations, or that its normative force may be amplified or mitigated by agent-relative considerations, and so on... but they are not likely to dispute that *other things being equal*, pain is rightly considered an intrinsic moral evil. It’s a controversial claim to hold pleasure as the *only* thing that has intrinsic value, but not a fraction as controversial as to deny that it is *one* of the things that do so. And as for the system’s tendency to place an uncommonly high premium on theoretical parsimony, this is censored for being *taken to excess* rather than for being viewed as methodologically desirable at all. Generally, the complaint against utilitarians is not that they are in error about their premises, but rather that they are in error for thinking their premises exhaust the full range of considerations with *sui generis* moral relevance. And if it *is* the case that these other considerations are intuitive *in the same way* that the utilitarian premises are intuitive, then it *would* be ill warranted to deny the legitimacy of such a complaint.
Section 3.2 – The “deontological” intuitions

3.2.0 – How do “deontological” or commonsense moral intuitions differ?

When the intuitions in this subsection are described as “deontological” in quotation marks, this is to reflect that – strictly speaking – it is more accurate to describe them as “integral to commonsense morality and notably at odds with utilitarianism.” But that’s a mouthful, and since these intuitions are being examined in the particular context of “intuitions that utilitarians are often criticized by deontologists for not accommodating,” simply referring to them as “deontological” seems an acceptable concession to convenience. Later in this chapter, we shall introduce an alternative terminology anyway. That being said, there is certainly no necessary link between deontology and commonsense intuitions. It is rather that this link, though merely contingent, is the link that is most interesting as far as this thesis is concerned.

There is another thing that had best be mentioned right on the outset: we shall not be concerned with the Categorical Imperative. This is not to say that it’s devoid of interest or anything, but – as far as this text goes – it’s not really an ideal subject for being investigated in this particular context and with these particular lenses. When we are talking about “deontological intuitions” here, we aren’t primarily concerned with abstract principles, although – and this bears mentioning – some of the intuitions in this section do overlap with formal principles. What we are mainly concerned with is what one might call lower-level phenomena. That is to say: we aren’t so much interested in overarching principles as we are in the kind of intuitions that many principles attempt to make sense out of, or – as is often the case with the Principle of Utility – are criticized for being too much at odds with.

Anyway: let’s cut to the chase. What are some good examples of intuitions that – for our purposes – could be considered “typically deontological?” Perhaps the first one coming to mind would be the intuition underlying the Doctrine of Double Effect. As with the Principle of Utility, the Doctrine is not itself an intuition. Intuitions and principles are two different categories, although principles that are found intuitively compelling will usually be so because they give propositional form to content that is implicitly entailed by one or more moral intuitions. As for this particular underlying intuition, we may simply refer to it as double effect and describe it as

---

The intermittent failure to emphasize this nuance may have contributed to the cavalier air with which some dismiss the principle as empty sophistry (Baggini and Fosl, 2007, p.133).

---
the intuitive tendency to judge that “even if a greater good is insufficient to justify a certain bad outcome as a direct and causally necessary precondition for it to obtain, the same greater good may yet suffice to justify the same bad outcome if the latter only obtains as an unavoidable but non-causative\textsuperscript{13} side effect of a permissible act that serves as the necessary precondition for the greater good to obtain.” When stated in such technical terms, this intuition may not seem immediately intuitive. Someone not already familiar with the principle may feel the need to mull it over a little before he is even confident that he has a grip of what it entails. But as soon as he is made aware that this is the intuition at work when we judge that “although it is not permissible to kill civilians as a direct means to end a war, it may yet be permissible to bomb a certain munitions factory as a direct means to end a war even if we foresee that this act will result in the death of civilians as an unavoidable side effect of destroying the factory” he is likely to agree that now it seems quite intuitive. So it turned out that the principle was intuitive after all: the initial stumbling block was rather that it was not obvious. And this is not disqualifying, for although the terms “obvious” and “self-evident” are sometimes used interchangeably in everyday language, it is not controversial to say that the two qualities are separate and that the one need not presuppose the other.\textsuperscript{14}

Compared to many of the other intuitions in this section, double effect is characterized by a high degree of “philosophical respectability.” That is to say: whereas many of the more inchoate case intuitions can be hard to “translate” into formal principles without seeming abstractly counterintuitive, this one has a fairly high degree of accept as a reasonable and philosophically legitimate principle (Edmonds, 2014, pp.28-31). When I describe an intuition as “not translating well,” this is to say that when someone tries to give an explicit verbal account of precisely what the intuition entails, it is difficult to formulate this in a way that seems plausible or theoretically satisfactory. I briefly foreshadowed this back in 1.2.4 – back when I was outlining those seven success criteria – by mentioning that I would use this term in reference to an intuition’s ability to satisfy the sixth criterion: abstract plausibility. Here’s the gist: when we talk about “translatability,” the intended meaning is the relative degree to which the implicit rule of a pattern of normative judgment (a. or c.) will seem intuitive if it is outlined in declarative terms and evaluated as a general notion (b. or d.). Now, the intuition we called double effect is an example of an intuition that “translates” pretty well. Or at least it does so when compared to many of the other intuitions we’ll be examining here. Even people who disagree with the

\textsuperscript{13} That is to say: the intended outcome would obtain even if the merely foreseen outcome did not.

\textsuperscript{14} Most mathematical proofs are self-evident without being obvious.
Doctrine are unlikely to view it as a complete non-sequitur that makes no sense in the context of ethical theory. It is also a notably relevant principle: one that utilitarians and deontologists frequently disagree about in practice. By deontologists, it is often regarded as a crucial constraint against the kind of hyperteleological ruthlessness they fear may be legitimized by utilitarian theory. In contrast, many utilitarians see the principle as a classic example of the kind of formalistic sophistry that is all too often used as a moral fig leaf when carrying out policies that could not have been plausibly justified on the basis of their outcomes alone.

Most of the things said about double effect will hold true for the intuition underlying another formal Doctrine as well: that of Doing and Allowing. One might dub this intuition acts/omissions asymmetry, and encapsulate its content along the lines of “greater normative force adheres to an act than to an omission, even if the resulting end states should be equivalent.” As commonsense moral intuitions go, double effect and acts/omissions asymmetry are unusual for being mirrored quite directly by formal, explicitly stated principles. One might speculate that they “translate” particularly well because they are sensitive to differences in causal structure. Given that the very concept of morality is so integrally linked to that of agency, it is not surprising that it seems plausible even in the abstract to ascribe independent moral relevance to an act’s precise mode of agency: directly intended as opposed to obliquely intended, or committed as opposed to omitted. As we shall see below, however, there are many intuitions in the “deontological” class that have a hard time making the jump from “an implicit rule of a judgment that seems compelling then and there” (a. or c.) to “an intuitively-plausible notion that conveys in declarative terms what said implicit rule entails” (b. or d.) without something seeming to get lost in translation.

3.2.1 – The ones that keep getting lost in translation

Whereas the Twin Doctrines in the previous subsection “translated” tolerably well, this cannot be said for many of the other intuitions we here call “deontological.” The thought experiment – due to Peter Singer – usually called “Shallow Pond” may serve to illustrate this phenomenon. If someone wearing an expensive suit were to witness a child about to drown in a shallow pond, as good as everyone will agree that it is morally obligatory to wade right into the pond and save the child, no matter the magnitude of the laundry bill this would entail. In contrast with this, we are not likely to find it even a fraction as despicable if someone omits to donate a sum – equivalent to what it would cost to clean or replace the suit in the previous example – to some effective charity that would then have used it to save the life of a child somewhere in a country far away. From a very abstract and technical point of view, we have two cases where an
agent omits to prevent someone’s death because doing so would incur a monetary cost. As long as they’re evaluated from that very distant perspective, one might have thought the cases were morally equivalent. But we know perfectly well that this is grossly at odds with how anyone would evaluate them. The omission is considered monstrous in the first case, and rather trivial in the second (Singer 1972). So far, one might have thought this was another of those anti-utilitarian thought experiments alluded to in 2.0.0. But as it happens, this is not the case.

Although we intuit clearly that there is a very significant moral difference between the two cases, it can be challenging to point out a morally relevant difference that seems plausible when stated explicitly. Since Singer is not sympathetic to the intuition in question, he suggests that the crucial difference may simply be that of geographic distance. And when evaluated purely as a descriptive fact about human psychology, this seems like a plausible explanation for why the two cases are judged so differently. But this psychologically plausible difference doesn’t seem at all plausible as a morally relevant distinction. Someone who held that we are not misguided when we intuit in accordance with this pattern – we might call it charity begins at home – would certainly not propose as a moral principle that “moral responsibility decreases in proportion with spatial distance.” He might concede that this described an overlapping psychological mechanism that was also at work, but he would not agree that it described the normatively relevant factor that our moral intuitions were sensitive to. But to point out that other, presumably relevant factor is challenging enough that “Shallow Pond” has become one of the keystone cases of what is sometimes called the Problem of Distance in Morality (Kamm 2007, p.345). Put simply: our case judgments strongly indicate that our intuitions treat distance as a morally relevant variable, whereas our tendency on abstract reflection is to intuit that there’s something just not right about this. Distance certainly seems to factor into our moral judgments, and it could be indirectly relevant in light of how it relates to some other consideration – but to say that it’s morally relevant in itself rings a false note. If there really is a morally valid reason for our differing judgments about these cases, it appears to be one that is difficult to “translate” into general and declarative terms.

Both cases involve omissions rather than acts, so one cannot invoke that distinction. It may be pointed out that in the charity case, there are a great number of other people who might produce the same good outcome as the agent, whereas in shallow pond it will either be done by the agent or not done at all. That, I shall grant, isn’t too unreasonable a suggestion. Nonetheless, it might give pause if someone were to suggest that the implicit rule of such an intuition could be summarized as “a blameworthy act or omission can only yield a limited total amount of blame,
wherefore it follows that the blame apportioned to each individual malefactor is decreased in proportion with the total number of malefactors implicated (\textit{diffusion of responsibility}). Once again, this sounded a lot more plausible as a description of a non-moral, psychological phenomenon than as a proposed moral principle. And although one may object that \textit{this} particular phrasing involved an element of epistemic rigging,\textsuperscript{15} the proposed account suffers from problems at case level as well. Even if there had been a lot of other people around, one still intuits that it would be horribly immoral not to save the drowning child. A similar problem occurs if someone should invoke shared nationality – one of many manifestations of the intuition we may call \textit{special obligations} – as a rationale for the difference in evaluation. We wouldn’t find it acceptable to let a nearby child drown if it had been a foreign citizen, so once again it seems hard to escape that \textit{spatial proximity} appears to be the most constant and invariable predictor for when this kind of omission is judged impermissible.

This could be continued for a long time, but the salient point has been illustrated: these intuitions seemed to suffer from a lack of \textit{“translatability.”} They seemed really compelling as long as they were embedded in a specific context, but had a tendency to seem off or not right when it was attempted to encapsulate in explicit terms what they implied on a more general, context-independent level.

And the intuitions above didn’t even \textit{“translate” that} poorly. For someone wishing to demonstrate the pitfalls of too much trust in generalizations from case intuitions, it is quite easy to formulate \textit{“intuitions”} (in the sense of implicit rules observable within commonsense moral judgment) that appear \textit{“untranslatable”} in the extreme. For instance, commonsense morality \textit{as manifested in practice} seems highly consonant with such implicit rules as \textit{wrath makes wrong} (“an act’s moral blameworthiness is directly proportional to the amount of anger it arouses in me”), \textit{yuck makes wrong} (same phrasing as previous intuition, but replace “anger” with “disgust”), \textit{only cruel when unusual} (“an act’s degree of moral blameworthiness decreases in proportion with the number of times the act is performed”), and the intuition we may call \textit{handsome does as handsome is or every man gets the face he deserves} (“moral blameworthiness is inversely proportional with physical attractiveness”). One is unlikely to find

\textsuperscript{15} That is to say, I formulated it in terms that highlighted its least-plausible aspects. But although this wouldn’t do if the example’s function was more profound than simply illustrating a point in passing, I also think such purposefully-uncharitable formulations may be less out-of-order than it would initially seem. Trying out different declarative interpretations is an important part of testing an intuition’s abstract plausibility, and the question of how implausible it \textit{could} be made out to seem is not, I believe, an insignificant dimension in this regard. In a sense, one may view this \textit{“counterargument by implausible formulation”} maneuver as the abstract-favoring flip side of the \textit{“counterargument by implausible implication”}-approach so prevalent to the Method of Cases.
ethicists who would take *these* implicit rules seriously as potential moral principles. But why not? As can be attested by countless jury verdicts, these wrong-on-their-face “intuitions” encapsulate persistent and highly influential tendencies in *that very* shared instinct or moral *sensus communis* so often valorized as the yardstick from which normative systems should deviate but subtly lest they be written off as pedantic *Hirngespinnst*.

That being said, we may not want to place too much emphasis on *that* particular batch of intuitions, seeing as an intuition-friendly ethicist could justifiably suggest that *these* examples demonstrate our capacity to use commonsense morality as our basis *without* sliding into excessive credulity. Nonetheless, they may open him to a charge of inconsistency somewhat parallel to the Only Game in Town: if one permits case-level intuitiveness to compensate for a lack of abstract intuitiveness in the case of *some* intuitions, why not others as well? It seems that if a generalization from case intuitions is found *actively unreasonable* in the abstract, it doesn’t really matter how compelling those case intuitions seemed down on the ground. If that is so, it is not immediately obvious that they should count for any more in cases where their implicit rule merely seems *mildly* unconvincing on abstract reflection.

3.2.2 – *It’s time to get more systematic*

So far, we’ve been taking a somewhat haphazard approach towards the intuitions in this section – beginning with the two Doctrines because that seemed like a very natural starting point, and then continuing to look at a brace of intuitions that were at the opposite end of the “translatability” scale. But now that we have examined both extremes of this important dimension, it’s best to get a bit more systematic in our approach. Seeing as the intuitions here in section 3.2 are all characterized by being at odds with classical utilitarianism, it is natural to structure one’s investigations along a few broadly defined classes – I call them clusters because they are rough tendencies rather than very strict categories – that I have dubbed “axiological non-isometry,” “axiological pluralism” and “praxeological precedence.”

**First tendency: The axiological non-isometry cluster.**

The intuition we called *acts/omissions asymmetry* belongs within this cluster, along with quite a few of the “untranslatable” ones. The intuitions in this cluster can be described as tendencies to reject what we may call *axiological isometry* – by that, I mean they run counter to the assumption that axiological considerations should always be tallied up with the *same scale of measurement* when assessing how much normative force they exert. So this cluster encompasses tendencies to judge that if a certain condition is met, we are justified in letting the magnitude of
some pains or pleasures (or other axiological markers, as applicable) count for more or less than they otherwise would. Two outcomes may be identical in the sense of involving \( n \) units of harm, yet we still find it compelling that harms that are caused count for more than harms one merely refrains from preventing. The harm may be equally doleful in both cases, but we still intuit that we have a greater duty to not cause it than to prevent it from happening through some other agency than ours.

Aside from acts/omissions asymmetry, what other intuitions can we identify within this cluster? There is one we alluded to already in 3.1.0, when remarking that it seemed counterintuitive to let the badness of pain and the goodness of pleasure count for precisely the same. This intuition may be dubbed value/disvalue asymmetry, and described as the tendency to intuit that – quantities being equal – the injunction to prevent or refrain from causing things that are bad (like pain) carries greater normative force than that of promoting or refraining from obstructing things that are good (like pleasure). Then we have the intuition we may call loss/no-gain asymmetry, and describe as the tendency to judge that it counts for more if someone loses something they already had than if they merely do not get something they didn’t have in the first place. Getting a $100 tax rebate isn’t anywhere near as fun as it’s vexing to pay $100 in back taxes, so this intuition certainly meshes well with our everyday phenomenology. Then we have yet another asymmetry that we may call self/other asymmetry (modest form), and describe as our tendency to judge that that the things we do to ourselves do not count for as much, morally speaking, as the things we do to others. Some may go further yet and find it most intuitive that the things we do to ourselves are outside the jurisdiction of morality entirely, but in that case they are intuiting the (strong form), which belongs in cluster three.

Not every intuition in this cluster includes the word “asymmetry” in its name, seeing as this is also the natural habitat of special obligations. This can be described as the intuitive conviction that there are certain agent-relative considerations (affection, consanguinity, promises, professional obligations, etc.) that give valid moral justification for acting in ways that would not be justified on axiological grounds alone. Even if the money used to buy a birthday gift for our mother could have produced \( n+1 \) rather than merely \( n \) units of happiness if we had donated it to charity, we are nonetheless adamant that she will be getting a present this year too. As mentioned above, those hard-to-“translate” intuitions in 3.2.1 – charity begins at home, diffusion of responsibility, only cruel when unusual, and so on – also belong in this cluster.
One may describe this cluster as the one least at odds with classical utilitarianism, for it can coexist not only with a moral system that is purely about outcomes, but also purely about outcomes evaluated in accordance with a monistic and cardinally-ranked axiology. They are merely intuitions about the parameters we are permitted to take into consideration when assessing how much – morally speaking – an outcome should count for. Intuitions like these will sometimes be accommodated even by theories that do not view them as normatively relevant at the basic level. For instance, we can see Bentham’s Disappointment-Prevention Principle as an auxiliary principle that takes loss/no-gain asymmetry into account, albeit for contingent reasons rather than as a premise that is theoretically basic (Crimmins 2015).16

Second tendency: The axiological pluralism cluster.

This category covers those intuitions that can still fit within a purely outcome-focused morality, but not one that operates with a monistic axiology. That is to say: these are intuitions to the effect that not all sources of value are reducible to (or subsumable under) a single dimension like pain/pleasure or preference satisfaction/obstruction, but rather that there are many different things that have value or disvalue in themselves. Along with this, it also covers intuitions to the effect that the sources of intrinsic value or disvalue aren’t uniform in the sense of possessing such valence in all circumstances. We see this cluster manifested when we intuit that the considerations that give us independent moral reason to promote or prevent something are irreducibly plural and not something that can be understood in terms of one concept alone (sovereign masters dethroned), when we intuit that there is a division between “normal” happiness that is good and “bad” happiness that is not (no rest for the wicked), or mutatis mutandis for suffering (retribution).

Third tendency: The praxeological precedence cluster.

Finally, there is a third cluster of intuitions indicating that under some circumstances, axiology is just plain beside the point. We intuit that there are some acts that won’t be acceptable no matter the outcomes aimed at (malum in se), there are some outcomes that are impermissible to produce as a means to an end while still permissible to produce as a foreseen but non-

16 That is to say, Bentham doesn’t hold that there is a fundamental difference between deprived and merely withheld benefit, but he holds that – due to contingent facts about human psychology and society – the deprivation of a benefit one had expected to retain is nearly always likely to result in greater disutility than the mere absence of an equivalent benefit of which there was no expectation. Therefore, it functions better to draft laws and regulate one’s conduct as if the DPP had been accepted as a properly basic principle. One may liken this to how a mathematician will employ proven theorems as if they had been proper axioms, simply because it would be an unnecessary waste of time to demonstrate their proofs anew every time.
causative corollary of pursuing an end by permissible means (our old acquaintance double effect), and if we consider it above-and-beyond the call of duty to produce a certain benefit or avert a certain harm, we accept no blame for omitting it no matter its magnitude to those concerned (supererogation). When intuitions like these are elicited, we simply stop counting because we intuit that there are some strictures on what is right that cannot be overruled by concerns for what will bring about the good. It is first within this cluster we find the constraints and prerogatives that characterize deontology proper as opposed to merely less uncompromising forms of consequentialism.17

3.2.3 – Analyzing the general differences so far

It seems the title of this chapter is doubly apt, for an insistence on distinctions appeared to be the guiding commonality that characterized the second class of intuitions. Rather than being directly at odds with the intuitions in 3.1, they were rather marked by a tendency to make us judge that under many circumstances, it’s just not as simple as it’s made out to be by those premises alone. Pain and pleasure count, but not always for the same. Many things derive their value or disvalue from their impact on our happiness, but they sometimes do so for other reasons as well. The outcomes aimed at are normatively relevant, but the same can be said for other considerations. Intended/foreseen, commission/omission, value/disvalue, loss/no-gain, self/other, distant/near, duty/supererogation – it seemed that the prevailing tendency of the characteristically “deontological” intuitions was an insistence that certain distinctions are normatively relevant in a way that cannot be accounted for in terms of utility alone.

Now, this invokes another echo from 2.2.0’s ruminations on the tendency for more-concrete intuitions to operate in an “apples and oranges” mode while their abstract counterparts favor an “all fruit” interpretation. It seemed that the intuitions in 3.2 were considerably more attuned to the “apples and oranges” perspective than their rivals in 3.1, which may cast light on our hunch about the “deontological” intuitions being abstract to a lesser degree than the “utilitarian” ones. As we recall from Chapter 2, there is a tendency for absolute boundaries to seem less intuitive in accordance with how abstract a perspective they are pondered from, and this is something that can quickly be noticed as a recurring trend among nearly all of these intuition, although the ones in the first cluster especially. It was mentioned already in 3.1.0 that

17 A category which, it should be added, includes versions that accommodate intuitions in all three clusters. Ideal utilitarianism accepts sovereign masters dethroned (second cluster), Mill’s liberty principle accommodates the strong form of self/other asymmetry (third cluster), while satisficing utilitarianism accepts a form of supererogation patterned in accordance with loss/no-gain asymmetry (obligatory to not make the world worse, supererogatory to make it better) (combines first and third cluster).
the intuition we would eventually call value/disvalue asymmetry exhibited the kind of problem one might call the vague/arbitrary gridlock: it seemed really intuitive that pain counted for more than pleasure, but it didn’t at all seem intuitive to suggest some particular proportion like “10 units of pain to 1 unit of pleasure” or anything such. And this is a dynamic that can be observed especially across the first cluster, seeing as it never seems quite right to postulate one particular number when outlining how much more actions should count for than omissions, or losses as compared to no-gains, and so on. It seemed that they needed to be kept vague in order to not seem counterintuitively arbitrary. This brings to mind the third criterion we presented in 1.2.4 – precision/clarity – so we’ll definitely have to return to this issue once we get to our normative comparison in Chapter 4. Right now, however, we are merely doing a descriptive comparison so as to demarcate the categories of our new taxonomy.

In what other ways did these intuitions differ from the utilitarian ones? For one, it seemed like some of them differed in mode of elicitation. We recall how double effect’s intuitiveness often needs to be “triggered” in a particular context before becoming apprehensible in the abstract. Recalling the methodological analogies at the tail end of 1.2.2, it may be interesting to note that this need for contextual elicitation is one that double effect shares with many linguistic intuitions. If a non-linguist is asked whether he finds it intuitive that size-predicates shall always precede color-predicates, he is unlikely to answer in the affirmative when the question is presented like that. But if someone asks “is it correct to say ‘black, big dog’ or ‘big, black dog’?”, then he will intuit that not only is it “big, black dog” but also that this ordering is how it must always be.

This seems promising if we want to outline a characteristic that makes it meaningful to call some general notions more “abstractly intuitive” than others. But is this something that applies widely to this class? Not in its most obvious sense, because most of the intuitions listed in 3.2.2 – the ones presented in connection with the clusters – are not at all likely to be found implausible when presented outside of any context. Nevertheless, there is a subtler sense in which also these intuitions can be said to have a less abstract mode of elicitation. Even if most of them are eminently commonsensical and easy to accept, they also have a tendency to seem considerably more intuitive when elicited in particular contexts than when contemplated in the abstract. Acts/omissions asymmetry, for instance, doesn’t seem entirely unproblematic from an abstract point of view – for one, there is that vague/arbitrary gridlock we mentioned, and then there is the more general tendency for the abstract mind to be wary of conflating the contingent with the noncontingent. But even if someone should come to doubt this intuition on abstract
reflection, he would still find it impossible not to think it worse to steal $100 from a charity collection box than to omit making a $100 donation to the same. When imagining such a situation, his intuitions bounce right back in place and he has to concede to himself that he cannot truly bring himself to believe there is no fundamental moral difference between acts and omissions.\(^\text{18}\)

There are two different ways of interpreting this phenomenon, which we shall refer to by the term abstract/concrete discrepancy. The first is to interpret it as analogous with the predicate-ordering intuition. If so, one may simply view the intuition as suffering from a lack of obviousness: it represents an implicit rule that is difficult to apprehend unless demonstrated in an appropriate context, but still eminently valid and worthy of consideration. If this analogy holds, then there is nothing epistemically derogatory about an intuition’s displaying concrete/abstract discrepancy.

But that first interpretation is not the one I will be arguing for. Rather, I shall hold that when intuitions display this discrepancy, we should take it to be an indicator that the abstract intuitiveness of their implicit rule is partly – perhaps even wholly – borrowed from the particular contexts in which it seemed intuitive. And this is problematic, because particular contexts often involve a great many distorting factors that we would – on abstract reflection – not regard as relevant. That is why so many contextually-intuitive judgments fail to “translate.” They fail to “translate” because our minds aren’t as easily fooled in the abstract mode as in the concrete. But of course, this corrective function is of limited use if we are still going to trust our concrete intuitions anyway, on the assumption that there’s no way they could be that compelling unless they were sensitive to some valid and reasonable implicit rule that we just haven’t “translated” correctly yet. If we find that certain notions keep growing less intuitive whenever we take a more abstract perspective, perhaps we should take that as an indicator of our abstract faculties trying to tell us how dubious this contextually-intuitive notion actually is? If we disregard these warnings, however, the inexorable forces of cognitive dissonance and implicit learning will do their work and gradually make us feel that these notions do after all seem kind of intuitive even in the abstract. And thereby – when this happens – is born what I shall dub a meso-intuition. Of which we shall hear more in the upcoming section.

\(^{18}\) Or so he cannot – at any rate – until the next time he switches into the abstract mode and asks himself “but was it truly the causal structure that made the crucial difference here, or was it rather…?”
Section 3.3 – There’s abstract and there’s “abstracter”: the tripartite taxonomy of moral intuitions

At this point, we may return to the questions raised in 1.2.3. Is Singer correct when he says that the intuitions he relies on are “different” from the ones he distrusts? And are they different in a way that’s methodologically relevant? When Kamm disputes that intuitions about cases are less reliable than intuitions about principles, are there any considerations she might recognize as valid counterarguments? Does it make sense to describe some intuitions as “rational” in a way that others aren’t? The stated goal of this chapter was to outline a methodologically useful distinction within the class of abstract intuitions. If one defines abstraction as a difference in kind, then every intuition not about a particular case is abstract. But we mentioned in 2.2.0 that abstraction may be more usefully seen as a difference in degree. And in that case, it may be relevant to distinguish between the upper and lower reaches of abstraction.

We started out with the hunch that the relevant difference may indeed be related to relative abstraction. But we needed to specify what this is supposed to mean, and why it should be relevant as a reason for preferring one class above the other. We shall now be doing the former task, while the next chapter will get into the latter.

3.3.0 – All good things are three

Here is the core of the position I advance: when we speak of case intuitions and abstract intuitions (whether about principles, general notions or theoretical propriety), it would profit us to operate with a distinction within the abstract category. There are some notions that seem intuitively plausible in a way that is more independent from case judgments than others. This isn’t to say that intuitiveness when applied at case level is unimportant, because it’s hard to see how a moral premise could make sense without a meaningful connection to the particular judgments we make. Nonetheless, I will hold that it makes a significant methodological difference how much its perceived intuitiveness depends on this connection. Compared to the intuitions in 3.1, the intuitions in 3.2 had a tendency to display comparatively less of this context-independence. They were abstract, all right – but they did seem to connect more with cases than the former class. Although the exact degree of concrete/abstract discrepancy varies across the different intuitions, all of them can to some extent be said to grow less intuitive the more abstract the perspective they are examined from. This matter – the question of whether or not an intuition seems to grow more and less compelling in accordance with how contextualized
it is – is the determinant of whether it classifies in the most-abstract or merely medium-abstract class.

The terminology around intuitions is often dichotomous, focusing on such concerns as the relative balancing of case intuitions against principles (as in the case of Kamm’s misgivings with Singer’s methodology), or in terms of balancing intuitions against desirable theoretical qualities (as is common in the context of RE). It is common to distinguish between case judgments (what is typically meant by “intuition”) and plausible general premises as a uniform class, but not so common to make distinctions within that latter class. I will argue that a lack of case-independent intuitiveness may count more strongly against a premise than what is usually held. Therefore, I will argue that one ought to distinguish between (1) case intuitions, (2) general intuitions with comparatively low case-independent intuitiveness, and (3) general intuitions with comparatively high case-independent intuitiveness. Since these distinctions relate to an intuition’s relative placement on the abstraction spectrum (particular-somewhat general-very general), I will apply the prefixes micro-, meso-, and macro- respectively: seeing as abstraction may, in a sense, be understood as a question of how wide a range of contexts something applies to.19

3.3.1 – The tripartite taxonomy

Our first category is the one we call **micro-intuitions**. These are intuitions about one particular context, whether case intuitions (a.) or methodological intuitions (c.). For practical purposes, however, they will by and large be used in reference to case judgments in this text. We needn’t worry too much about this class: it’s merely included for completeness, and – as far as case judgments are concerned – do not differ from the commonly accepted notion.20

19 I must confess to a slight feeling of linguistic unease when applying prefixes like micro- and macro- to abstract classifications of mental events (and their implicit rules) that have no spatial dimensions. However: when introducing new terminology it is always a good idea to use designations that have no previous history within the field. It helps to avoid confusion. That being said, the classes called micro- and macro-intuitions can equally well be referred to as simply “case judgments” and “abstractly cogent premises” respectively. It is mainly the intermediate class – meso-intuitions – that has no straightforward analogue within the established terminology. And as luck has it, this latter classification also happens to seem less misapplied than the other two. Whereas the prefixes macro- and micro- are associated very closely with size in the strictly spatial sense, the meso- prefix tends to be associated with a more general and non-specific conception of intermediacy (can be temporal as in Mesozoic, geographic as in Mesopotamia, anatomic as in mesencephalon, etc.). It is fortunate that this should be so, as it’s mainly in reference to this particular intuition class that it’s necessary to use this idiosyncratic terminology. As for the other two classes, I may at times permit myself to slip into less peculiar-sounding synonyms. It is mainly in the case of meso-intuitions that a new terminology is strictly necessary, whereas the other two classes are given alternate designations mainly for reasons of terminological symmetry.

20 As for methodological intuitions, it would be a bit hasty to refer to a “commonly accepted notion” at all, seeing as this concept has – to my knowledge – not really been subject to much systematic enquiry at all. When I include them in the micro-intuitive class, it is because they too are judgments about one particular context, such
Our second category is the one we call **meso-intuitions**. There may not be a very clear analogue to this class, although the term “principles derived by induction from (case) intuition” is a good fit. We briefly mentioned this class at the end of the previous section: they are the ones that – on my interpretation – our abstract intuitions are warning us about with rather modest success. Their defining feature is that they display significant discrepancies between case-level and abstract intuitiveness. If an intuition is such that it never seems quite satisfactory on abstract reflection, but then grows a lot more compelling again when elicited in an appropriate context, then it is a meso-intuition. The term “significant discrepancy” is necessarily vague. In this paper, I will argue for treating pretty much any discrepancy as “significant,” but I won’t insist that this is the only valid interpretation.\(^{21}\) However, as far as this text goes, the whole class of intuitions described in 3.2 will be subsumed under this class. It includes both the implicit rules observable within patterns of case judgments (a.) and general notions that seem intuitive (b.), overlapping to some degree. However: they will necessarily be experienced as *more* intuitive when manifested as (a.) than as (b.) – seeing as they wouldn’t *be* meso-intuitions without such a discrepancy.

Although I will not *preclude* the possibility that this class could include methodological intuitions (c.), I hold it to be fairly unlikely, seeing as methodological intuitions tend to be intuitive in a distinctly abstract way. Due to the abstract preference of methodological intuitions, I will tentatively assume that – save for the possible exception of **double effect** – this class does not include intuitions of the (d.) category.

Our third category is the one we call **macro-intuitions**. Terms like “rational intuitions,” “properly self-evident premises” and “Axioms of Reason” may be seen as applicable to this class. The defining feature of this class is that there is no significant discrepancy between concrete and abstract intuitiveness involved – or at least not in the concrete-to-abstract direction, at any rate. These are intuitions that seem highly plausible in the abstract and on the outset, although they can still lead to implausible implications when applied to particular cases. Because these intuitions are so abstract and general, they tend to have a very wide scope, and it’s uncommon to find someone who will deny finding them compelling to some degree. In this category, we find the intuitions described in 3.1 – the ones we called “utilitarian.” This class encompasses intuitions in categories (a.) (implicit in case judgments), (b.) (implicit in

\(^{21}\) As indicated in 3.0.1, I do entertain a slight hope that this terminology could be found methodologically useful in a wider context, so I wouldn’t want to tie it too closely to my own substantive position on the relative merits of rivaling normative systems.
methodological judgments), and (d.) (general notions that are found intuitive and congenial to our methodological sensitivities).

3.3.2 – What now?

We have now arrived at a classificatory scheme that will allow us to speak meaningfully of “utilitarian” intuitions as being distinguished from the “deontological” ones by their greater degree of abstraction. It is therefore time to proceed to the next chapter, where we shall look into the considerations that could justify privileging one class above the other.
Chapter 4: Evaluations

– The chapter where I put the concepts from Chapter 1 in conjunction with those from Chapter 3 in order to build a case for the position I advanced in Chapter 2

4.0.0 – Chapter Introduction

It is now time to focus more directly on the working hypotheses and success criteria presented in 1.2.4. My first working hypothesis held that the methodological approach towards normative ethics should be one that – compared to extant practice – put less emphasis on descriptive success criteria as opposed to normative. My second working hypothesis held that one could profitably examine the approach of classical utilitarianism as an example of how to achieve a more appropriate balance between these criteria. Chapter 2 elaborated on the second hypothesis by postulating that an uncommonly high weighting of what seems abstractly intuitive may be a crucial ingredient behind the utilitarian theory’s past successes at staying in the forefront of moral progress. We did, however, run into an initial impasse upon finding it unclear how well this postulated abstraction-utilitarianism connection could be applied to the methodological landscape. Chapter 3 set out to address this problem, eventually arriving at a distinction between a more-abstract class of general intuitions we called macro-intuitions, and a less-abstract class we called meso-intuitions. Our examination was then focused upon finding some overarching feature that would allow us to distinguish between the two classes in a descriptive sense. In this chapter, we shall continue by outlining the methodologically and epistemically salient differences that could justify treating the former class as privileged to the latter.

In the upcoming sections, we shall examine the differences between macro- and meso-intuitions in order to see if we can rightly say that the former class compares favorably to the latter in accordance with the success criteria outlined in 1.2.4. First, section 4.1 will be looking into how the two intuition classes compare in terms of the methodological criteria we dubbed universality (3), precision/clarity (4), consistency (5), and abstract plausibility (6). Following this, section 4.2 will compare how well the meso-intuitions can retain their presumed baseline plausibility (2) in the face of attempted obviation. Finally, section 4.3 will sum up how the foregoing examinations can ground the claim that macro-intuitions are relevantly different from the other classes in ways that may deflect the Only Game in Town objection.
Section 4.1 – The methodological case for macro-intuitions

In this section, we shall compare how well the macro- and meso-intuitions compare as far as criteria 3-6 – universality, precision/clarity, consistency and abstract plausibility – are concerned. One may regard this section as comprising the pragmatic-methodological argument for privileging macro-intuitions.

4.1.0 – We return to the subject of “translatability”

It is natural to start with the abstract plausibility criterion because it connects so directly with the considerations examined in Chapter 3. When we addressed the phenomenon we called “translation” in that chapter, it was mainly as a route of approach for arriving at the descriptive characteristic we called abstract/concrete discrepancy. We noted that there was quite a gulf between the more “translatable” meso-intuitions – such as double effect and acts/omissions asymmetry – and those that were so hard to “translate” that it’s ambiguous whether they belong under the moniker meso-intuition at all.22 We did note already then that “translatability” was tightly connected with abstract plausibility, but postponed a closer examination of this relation until Chapter 4.

As was amply noticeable when comparing the intuitions in 3.2.0 and 3.2.1, the meso-intuitions can be divided along a pretty wide spectrum according to how poorly they rate according to this criterion. There were some who merely seemed to lose their intuitiveness in the abstract, some that started to seem problematic, and some that appeared objectionable in the extreme when “translated.” Double effect is an example of the first type: it doesn’t seem unreasonable or paradoxical in the abstract – it’s just that many will find it hard to see precisely why this difference in causal structure should make such a crucial moral difference. This may be described as a form of weak “untranslatability” – trying to “translate” the intuition merely makes it seem somewhat lacking in positive justification. In contrast, charity begins at home was hard to “translate” in a strong sense: although it seems conceivable that distance could be indirectly relevant in light of how it relates to some other consideration, we find it very implausible that distance in itself should be a morally relevant parameter. Every man gets the

22 If an intuition has no abstract plausibility at all, it would not belong in the hybrid category we call meso-intuitions. However, I avoid making a very definite judgment about the classification of the least-“translatable” intuitions since it’s not so sure everyone would view their implicit rules as completely devoid of abstract plausibility. Leon Kass’ notion of the “Wisdom of Repugnance,” for example, can be viewed as a “translation” of yuck makes wrong’s implicit rule that is found plausible by some (Herz 2012, pp.199-200). One may also – perhaps a little speculatively – view the long-discredited field of physiognomy as an attempt to provide a scientific fig leaf for judging in accordance with the every man gets the face he deserves-intuition, seeing as the kind of features that were regarded as indicative of a base or criminal nature had a suspicious tendency to coincide with those that were disfavored by the aesthetic ideals of the time. 

64
face he deserves seemed to display extreme “untranslatability.” Whereas it’s possible that some deeper-level principle could justify the indirect normative relevance of distance, it seems hard to imagine any kind of underlying principle that could justify a jury’s decision to acquit a felon due to the pleasuring way in which the angles of his cheekbones accentuated his bright blue eyes.

So although this abstract/concrete discrepancy in intuitiveness was very important as a general characteristic for distinguishing between macro- and meso-intuitions in a descriptive sense, it would beg the question to think this alone enough to consign the full range of meso-intuitions to the status of epistemic second-class citizens. Although such a discrepancy could be indicative of there not being anything to “translate” at all – i.e., the relevant case judgments came cloaked in the phenomenology of moral judgments while in reality being responsive to something morally irrelevant – it could equally well be that they responded to morally relevant parameters that were subtle and difficult to identify. It certainly wouldn’t mesh well with the philosophical outlook to shy away from premises that seem puzzling and in need of further investigation, so we can’t rely on this dimension alone when trying to rebut the Only Game in Town objection.

4.1.1 – Universality and precision/clarity

In a way, one might describe macro-intuitions as intuitions about the parameters that are morally relevant other things being equal, while meso-intuitions tend to be intuitions about the parameters that make other things legitimately unequal. At the end of 3.1.0, we noted that the utilitarian premises weren’t just accepted by utilitarians, but pretty much by everyone. Seeing as one could hardly portray a premise as lacking in universality if ethicists all across the board would regard it as a strawman portrayal if someone accused them of not taking it into account, it can safely be said that the macro-intuitions are doing very well by this criterion. But it’s not so obvious that the meso-intuitions are doing any worse. It is, after all, very common to intuit that other things aren’t always equal, and the meso-intuitions did seem to conform with very widespread notions about which parameters make things legitimately unequal.

Take supererogation. Although the term itself may be obscure to non-philosophers, its notional content – we may encapsulate it as “there are some acts that carry normative force in the sense that an agent merits praise for performing them, while not carrying normative force in the sense that an agent merits blame for omitting them” – cannot at all be described as marginal. Most people do find it plausible that it makes sense to differentiate between what duty requires
and that which goes above and beyond it. And this goes for a great many other meso-intuitions, such as special obligations (“there are certain agent-relative considerations that give valid moral justification for acting in ways that would not be justified on axiological grounds alone”), sovereign masters dethroned (“there are some concerns that carry intrinsic moral value/disvalue independently of their impact as far as hedonic axiology is concerned”), and malum in se (“there are some acts that are intrinsically impermissible, and this impermissibility is not subject to being outweighed by axiological considerations”). Arguably, it is far more common to accept these notions than to reject them, which is indeed why the utilitarian theory keeps getting on deep water as far as descriptive fidelity is concerned.

But despite this, there is a sense in which the meso-intuitions can be viewed as having considerably less universality than the macro-intuitions. All of the intuitions described above were formulated in a rather vague and nonspecific way. More to the point, they all included phrases like “some concerns,” “certain considerations,” and “some acts.” Those phrasings were too general to be applicable to any substantive moral judgments as they now stand. One would first have to specify which concerns, considerations or acts one were having in mind. Sovereign masters dethroned, for instance, rates very high in universality in the minimal sense of asserting that pain and pleasure are not the only things with intrinsic axiological valence. But it’s not so universal when it comes to the specifics. There are some who ascribe intrinsic value to beauty, while others do not. Some may hold that the only thing good in itself is the good will, whereas others again will ascribe intrinsic value to such qualities as knowledge, autonomy or personal growth. Although axiological pluralists share the conviction that there are other value premises that deserve the same degree of consideration as the hedonic ones, they are hardly unanimous about which value premises do. Special obligations runs into a similar

---

23 Supererogation was one of the main intuitions I had in mind when I described utilitarianism as taking “some pretty extensive liberties with our very basic moral notions” in 2.3.1.

24 Taking a more systematic approach, we find that early intuitionists like Clarke and Price ascribe intrinsic value to piety (one has a fundamental, non-derivative duty to honor God), whereas later intuitionists like Prichard and Ross do not. Clarke combines this with the basic duties of beneficence and equity (i.e., impartial concern for well-being, so that just underlines the universality of the macro-intuitions) and that of self-preservation (some would call this prudential rather than moral, so again not universal). Price, in contrast, displays an inegalitarian streak as he dispenses with equity while adding justice (a term that denoted “respect for property” at the time) and the duty to respect one’s superiors to the list of fundamental, nonderivative building blocks of morality. Some view the value of fidelity to promises as extrinsically grounded on the properly basic duty of veracity (Price), whereas others see the value of such fidelity as intrinsic in its own right (Ross). Sometimes, intuitionists do not even agree with themselves: Ross initially presents seven prima facie duties, but scales the list down to five after concluding that the duties of justice, beneficence and self-improvement may better be seen as derivative from the more basic duty to promote the good, leaving fidelity, gratitude, reparation, and non-maleficence. Given the variety observed among moral tenets whose non-derivative, theoretically basic validity
problem while also running afoul of the vague/arbitrary gridlock – whereas all will agree that an otherwise commendable compliance with this intuition can slide into blameworthy nepotism if carried too far, few can give a clear and consistent account of where that threshold lies. This tendency to run into the Scylla of the arbitrary whenever one tries to escape the Charybdis of vagueness is something that bedevils any kind of moral distinction that involves thresholds, which would include not just supererogation but pretty much any meso-intuitive constraint or prerogative that is recognized as potentially defeasible by axiological considerations if only the stakes get high enough. As for malum in se, this intuition may appear subject to greater unanimity as long as we’re concerned with especially odious acts like rape and torture. But as for those acts, it’s not really as if one would have to declare their unconditional impermissibility to be a theoretically basic premise, seeing as it could just as well be grounded on the ample indirect-utilitarian reasons for enforcing unconditional sanctions against them. And as for the notions of intrinsic impermissibility that cannot be grounded on such rationales, those aren’t very universal at all, but rather quite parochial and variable.

These problems most definitely connect with our third criterion: precision/clarity. It appears these meso-intuitions were subject to widespread and considerable certainty as long as the details were kept vague, but had a hard time retaining that quality while simultaneously pursuing that of precision/clarity. And this isn’t just a problem to those meso-intuitions, but rather one that keeps bedeviling nearly the whole class.25 As we briefly noted in 3.2.3, the phenomenon we have taken to calling the vague/arbitrary gridlock was especially endemic – but by no means exclusive – to cluster one. These intuitions were all concerned with some distinction or other that seemed plausible as long as one didn’t get into the specifics, only to be found unclear and contentious as soon as one did.

But are the macro-intuitions really doing so much better when it comes to precision/clarity? Granted, positive and negative hedonism do have problems in this department, seeing as they too run into the vague/arbitrary gridlock if one tries to quantify pain or pleasure. Nonetheless, they are still a great deal more clear and precise than other value premises in the sense that pain and pleasure are a lot less likely to be interpreted disparately than the various other qualities proposed as intrinsically valuable. People can have really differing ideas about what it means more specifically to respect the person, honor human dignity, let justice be done, was supposed to be self-evident, one may find it understandable that Prichard plays it safe by declining to even present a list (Stratton-Lake 2010, pp.470-474).

25 Double effect is the odd man out here, which is probably why it’s considered such an ideal candidate for a legitimate, formal principle. As we shall soon see, however, it has problems of its own.
or promote flourishing, but where pain and pleasure is concerned they are quite on the level. As for less is more and same rules apply, one might at first glance have thought these to be quite unclear, owing to the widely differing views on how much they should count for. Although it cannot be denied that these are problematic issues, the intuitions themselves can still be called eminently precise in their formal properties. They are intuitions to the effect that whenever the choice is between increasing or diminishing simplicity or consistency, the burden of justification always falls upon the one who opts for a diminution. And this – “if the choice is between more-x and less-x, the latter requires justification whereas the former does not” – is a very clear and precise rule. When the proponents of rivaling systems censor utilitarians for carrying the pursuit of simplicity or impartiality too far, it is not because they refuse to accept that the one who would sacrifice them bears the burden of proof, but rather because they are convinced that the burden of proof has been satisfied by pointing out relevant, countervailing considerations.

So the macro-intuitions can indeed be called more clear and precise, seeing as the notions of pain and pleasure leave a lot less room for interpretation or ambiguity than other proposed bearers of intrinsic value, and the lack of clarity surrounding less is more and same rules apply is not actually a lack of clarity about their implicit rules so much as a lack of clarity about which countervailing considerations can serve to justify a deviation from them.

4.1.2 – “Clear and certain” vs. “clear or certain”

Whereas the macro-intuitions could be called “propositions of real clearness and certainty” – a phrase used by Sidgwick to describe his Axioms of Reason – it seemed more apt to describe the meso-intuitions as propositions of real clearness or certainty, but not both at the same time. If a premise can only serve as a dialectical common ground as long as it’s kept too vague to derive substantive positions from, that certainly limits its usefulness.

Worse yet, this leads to the risk of conflating the unanimity directed at a vaguely defined premise with the far more limited agreement directed at one’s own interpretation of its precise details. For instance: since it is found so widely plausible that some (unspecified) things have intrinsic value independently of pain and pleasure, it is easy to conflate that consensus with the more modest consensus due to one’s own notions about specifically which things do, thus taking for granted that every right-thinking certainly person agrees that such-and-such a good is so uniquely valuable that it must be promoted even at the cost of considerable sacrifices in simple happiness. Needless to say, this is the kind of mix-up that can cause a great deal of mischief if it happens in the head of a legislator. And this tendency isn’t just a prime recipe for confusion, but
also for outright vehemence. Mere disagreement is one thing, after all, whereas disagreement that \textit{appears} to be a highly deviant failure to see what is mistakenly assumed to be the consensus of all reasonable persons is quite another.

So at this point we are arriving at more substantial reasons to be wary about extending too much trust in our meso-intuitions: they can easily exacerbate disagreements because they appear deceptively universal as long as they are cloaked in vagueness, while in actuality subject to a great deal of contention as soon as the details are filled in.\textsuperscript{26} In the interest of fairness, it has to be admitted that the macro-intuitions \textit{do} contain some vague elements themselves, so it’s not as if they can be viewed as \textit{entirely} immune to this problem. However, they \textit{do} have an advantage in that the parts that have to be filled in (how to quantify utility, etc.) are likely to be \textit{recognizable} as subjects for interpretation rather than mistaken for the common consensus of mankind, whereas the elements presumed universal arguably \textit{can} be described as such. The potential for mix-up doesn’t seem quite the same. More so, they do not \textit{lose} their other methodologically desirable qualities when made more precise.

\textbf{4.1.3 – Consistency}

In the foregoing subsections, we found that the meso-intuitions weren’t subject to as much intersubjective agreement as it initially seemed. And as it happens, they’re not always doing so well as far as \textit{intrasubjective} agreement is concerned, either. That is to say: not only do people disagree with \textit{each other} about these intuitions, but they aren’t always on the level with \textit{themselves} either.

The macro-intuitions tend to stay pretty constant. There are certainly \textit{some} cases in which they seem less intuitive: \textit{positive hedonism} doesn’t seem so intuitive in cases that elicit \textit{no rest for the wicked}, for instance. But even here, one is more likely to interpret their momentary counterintuitiveness as a result of a particular defeater condition being applicable, rather than as disproof against their validity \textit{ceteris paribus}. In contrast, there are many meso-intuitions that vary widely between different contexts. As we could see from \textit{Shallow Pond}, there are some cases where people do not care one whit about \textit{acts/omissions asymmetry}, and the intuitive appeal of \textit{double effect} varies considerably. It’s probably not a coincidence that the example with

\textsuperscript{26} This is especially so since moral intuitions – it may be argued – aren’t just experienced as self-evident in the \textit{descriptive} sense, but also as \textit{normatively self-evident}. That is to say: it’s not just that we \textit{do} find them self-evident, but also that we strongly feel they \textit{ought} to be found self-evident, and take a very dim view of those who seem not to. This, understandably, can become problematic if the intuition in question is \textit{actually} very interpretation-hungry, and leaves a lot of blank space for the individual to fill in the specifics.
“bombing civilians vs. bombing a munitions factory” is such a common choice for illustrating this intuition, because that particular case is one in which it seems highly plausible.

As for other cases, it’s rather hit-and-miss. Take the euthanasia loophole, for instance. Many jurisdictions that prohibit medical euthanasia de jure will still let it slide de facto if the lethal dose of anesthetic is administered with the express purpose of alleviating the patient’s pain while merely foreseeing that a dose of that magnitude will also result in death as a “side effect.” As much as people’s intuitions differ, it seems safe to say that it carries a whiff of magical thinking to ascribe a crucial moral difference to whether or not the act of knowingly terminating a patient’s life is accompanied with a ritual incantation-like insistence that pain relief is the only outcome being expressly intended despite being irrelevant in light of the presumed “side effect” that will accompany it (Baggini and Fosl 2007, pp.132-133).

And then there’s Hiroshima. That was pretty much a parade example of a failure to satisfy the requirements of double effect, but the opinions on its ultimate justifiability are pretty variable, and – tellingly – tend to focus on whether or not it was the best option available rather than its causal structure. In contrast, the firebombing of Dresden is condemned nearly universally for its gross disproportionality between casualties inflicted and strategic advantage gained. Both of these cases were similar in that they were recognized as in violation of double effect, whereas they were different in that Hiroshima might (or might not) have been justified by its ultimate outcomes while the bombing of Dresden is unambiguously regarded as instrumentally useless and unnecessary. This is the kind of contrast that may lead one to suspect the causal structure issue of being a red herring.

How endemic are these problems to the class of meso-intuitions at large? That may be hard to say, seeing as they are often so vaguely defined that it’s difficult to even measure the consistency of their applicability. Double effect and acts/omissions asymmetry are unusual for being among the few meso-intuitions that are clear and precise enough for cross-case testing like this, and they didn’t seem to do all that well.

4.1.4 – Connecting the threads

It quickly became evident that the four criteria of abstract plausibility, universality, precision/clarity and consistency were tightly interconnected with each other. Notions that seem plausible on abstract reflection and consistently applicable to our judgments are the ones that will be accepted near-universally even when outlined in clear and precise terms. If only one of
the links in this chain is found missing, the intuition in question will start to take on an element of abstract/concrete discrepancy that impacts its suitability as a premise in one way or another.

*Charity begins at home* is both consistent and universal, but difficult to outline clearly in a way that seems plausible on reflection without losing some of its consistency of applicability. *Acts/omissions asymmetry* and *double effect* are among the most “philosophically respectable” meso-intuitions because they are clear and precise enough to be turned into formal principles that are found widely plausible on reflection, but they didn’t actually seem to apply very consistently to cases. Most of the asymmetries in cluster one seem plausible and universal until one runs into the vague/arbitrary gridlock when trying to clarify them, and *sovereign masters dethroned* had to stay vague to retain its universality. And so it goes. Although the meso-intuitions varied widely in how badly they were doing, they all seemed to lack something or other.

And all along, the notions expressed by the *macro-intuitions* remained as a constant backdrop. To a great part, the meso-intuitions were defined by their relation to them. These may all have been intuitions to the effect that certain apples shouldn’t be mixed with certain oranges, but they all came bundled with the implicit understanding that it still mattered *ceteris paribus* that both apples and oranges are fruit. And about this *ceteris paribus* presumption, they seemed to display greater unanimity than they did about the various considerations that would sometimes overrule it. They were intuitions about the causal structures, agent relations, value premises, and so on that would sometimes make it right to deviate from what would otherwise be implied by the macro-intuitions. And everyone finds it plausible that it sometimes is right. The problem is that they only agree about this when it’s phrased in its vaguest and most general terms. Lots of people agree that other factors count, but they agree far less about precisely which factors do, and in what proportion, and in what way they relate to each other.27 It appears – to put it in the parlance of politics – that the opposition is split.

Now, could we turn this into a rebuttal of the Only Game in Town objection?

---

27 As an illustration of this latter point, we may observe that the intuitive appeal of *value/disvalue asymmetry* is nearly universal as long as it’s applied to the pleasure/pain dimension, but far less so when applied to other value premises. If one accepts a form of *sovereign masters dethroned* patterned along perfectionist lines (developing one’s capacities to the fullest has intrinsic value), there is little consensus as to what priority should be given to avoiding the disvalue of not being well-rounded as opposed to pursuing the value of developing a particular capacity to its greatest possible extent (Hurka 2006, pp.368-369). This tendency for additional premises to seem strongly plausible on their own yet problematic when combined – a philosophical analogue to what would be called *adverse cocktail effects* in medical contexts – is another, more subtle manifestation of less-than-stellar consistency.
Rebuttal #1: The opposition is split. Macro-intuitions tend to be accepted *ceteris paribus* and mainly subject to *indirect contention* in the sense that many other considerations (agent relations, causal structures, non-derivative value premises, various constraints and prerogatives, etc.) are ascribed countervailing force. Meso-intuitions are intuitions *about* those countervailing considerations, but although they *seem* very widespread and certain, this only holds true as long as they are kept too vague to be of much use as formal premises. When they are outlined in more precise terms, it turns out that they are actually subject to significant *direct* contention. This can be a valid reason to ascribe them less evidentiary weight than macro-intuitions for reasons that are both *pragmatic* (it is desirable to avoid relying on contentious premises) and *epistemic* (if there is disagreement, this presumably means at least *someone* is wrong, which implies that this is an intuition that it’s *possible* to get wrong, which in turn should affect one’s trust in one’s own ability to get it right).

How far can this rebuttal get us? Not as far as we’d like. These considerations *do* count for something, and they have their place in the overall case for favoring macro-intuitions. But a case relying on them alone may be too *circumstantial* for comfort. A lack of *universality* may indicate error-proneness and greater vulnerability to factors that aren’t found impartially valid, but it doesn’t *demonstrate* it. It may be a bad thing that an intuition is difficult to summarize in clear and explicit terms without losing some of its perceived certainty, but it doesn’t follow that we should stop trying. Taken on their own, these considerations could equally well imply that the macro-intuitions are intuitive in a more *obvious* and *accessible* way than meso-intuitions. That is to say: they are *obvious* because their intuitive appeal can usually be apprehended without having to consider particular cases, and *accessible* because it can be apprehended by nearly everybody. Although these qualities are relevant as a reason for favoring them in contexts where such accessibility is at a premium – policy deliberations in the public sphere, for instance – the same may not hold true for fundamental ethical theory. As it stands right now, this rebuttal looks a bit too much like an appeal to popularity to fit in a strictly epistemic context, so we had better continue to section 4.2 and see if we can shore it up with additional rationales.
Section 4.2 – The epistemic case against meso-intuitions

4.2.0 – The evidence is as yet too circumstantial

We ended the previous section by noting that although our first potential rebuttal of the Only Game in Town (R1: The opposition is split) could carry us a fairly long way if the context had been more strictly pragmatic (in which case consensus counts for a great deal), its epistemic aspect seemed too circumstantial to be satisfactory in this context. That is to say: we pointed out various methodological shortcomings (along criteria 3-6) that could indicate that the meso-intuitions were subject to error sources that the macro-intuitions avoided, but they didn’t indicate it strongly enough. However, such indicators of possible error-proneness can function well as a starting point for a more explicit account of what the error sources could consist in.

This brings us to the second criterion we outlined in 2.2.4: an intuition’s ability to retain its presumed baseline credibility in the face of attempted obviation. When I conceded that the rebuttal outlined in the previous section seemed a bit too weak to deflect the Only Game in Town objection on its own, this was precisely because the methodological shortcomings of the meso-intuitions didn’t on its own carry enough weight to counterbalance this “innocent until proven-guilty”-postulate. But if we can connect those shortcomings with some plausible account of why they keep having these problems, then we may have a stronger case.

What leads do we have to go by?

Lead #1: “different processes with different functions.” This phrase was used in 2.2.0, when describing the different tendencies of our concrete and abstract intuitions. It was implied that when we are having intuitions about decontextualized as opposed to context-embedded matters, this involves different mental processes that have been geared to carry out different functions. If we can make a credible case for thinking that meso-intuitions arise from processes that have been geared for different functions than the ones that serve us well in the context of formal ethical theorizing, then we would be having a stronger case for arguing that their methodological shortcomings are the product of more fundamental problems than a mere lack of correct application and understanding on our part.

Lead #2: “the intuitions to which Haidt and Greene refer.” This phrase appeared in the Singer quote in 1.2.3, where he held up this class as “different from” the intuitions that utilitarians rely on. It was evident from the context that he held it acceptable to trust “utilitarian” intuitions because they weren’t the type Haidt and Greene referred to.
These two leads seem promising. We might get somewhere if we combine them. A fuller and more satisfactory rebuttal to the Only Game in Town objection might be produced if we can make a case for the following:

a.) That meso-intuitions can be likened to Haidt/Greene-type intuitions while macro-intuitions cannot.

b.) That Haidt/Greene-type intuitions arise from “different processes with different functions” from macro-intuitions, and that an account of these “different processes with different functions” can obviate the presumed baseline credibility of the meso-intuitions so as to provide us with a fuller and more satisfactory rebuttal of the Only Game in Town objection than what R1: The opposition is split can provide on its own.

Trying to ground such a claim is the aim of this section.

Would it be appropriate to describe meso-intuitions as “the intuitions to which Haidt and Greene refer”? The answer is ambiguous, seeing as the bulk of their research is concerned with what Chapter 3 dubbed micro-intuitions. However, the entire purpose of that chapter was to get a clearer grasp of the relation between case intuitions and general intuitions, and then – in particular – the ways in which some general intuitions could be called less independent from cases than others. So this is a problem we are well equipped to address. If the meso-intuitions are “borrowing” part of their intuitiveness from the cases they are applicable to – and this is what we have taken the abstract/concrete discrepancy to imply – then it may well be that the problems that attach to those case intuitions may be problematic to the meso-intuitions doing the “borrowing” as well.

4.2.1 – So what kind of intuitions are Greene and Haidt referring to?

We mentioned in 1.1.1 that the boundaries between emotions and intuitions can be unclear, and that they are typically viewed as more porous by moral psychologists than by moral philosophers. It wasn’t always like that. At the time Jonathan Haidt began his studies in moral psychology – subsumed under the larger umbrella of developmental psychology at the time – the field was mainly focused on moral reasoning as one form of cognitive proficiency among others, directing its attention overwhelmingly towards abstract moral reasoning rather than emotion (Haidt 2012, pp.4-26). Haidt’s work in the field would soon come to challenge its rather optimistic view on the role of moral reasoning as the primary motive force of moral judgment, as he set out to demonstrate the phenomenon that would eventually be known as “moral dumbfounding.” The classic experiments involved presenting test subjects with a battery of
described situations that were apt to raise intensely negative emotions – people eating the family
dog for dinner, using a dead chicken for sexual gratification, using a discarded national flag for
cleaning the toilet, and so on – before eliciting their moral evaluation of the acts. Understandably
enough, the respondents condemned the acts nearly unanimously, but there was a snag: the
situation had been described in such a way that it was hard to point out any easy-to-accept
grounds of moral condemnation. 28 Thus, the test participants commonly found themselves at a
loss when asked to motivate their condemnation of the acts described. They would typically
begin by referring to some aspect that would commonly apply – under more normal parameters –
but were then reminded by the test administrator that such a factor had been specified to be absent here. Some respondents would accept, more or less begrudgingly, that although they
couldn’t help finding the act distasteful, it still might not be wrong in a strictly moral sense,
while others stuck to their guns and reiterated that the act was “just plain wrong” without giving
much of a rationale (Haidt 2012, pp.38-40). However: whether they changed their minds or not,
both groups seemed to have arrived at their judgments before they knew if they could be justified
or not, which was taken to imply that they had arisen in response to some other factor than an
awareness of parameters recognized as morally relevant. And as for what that “other factor” may
have been, few would disagree that emotion seems like a plausible candidate. A lot of these cases
aroused the emotion of disgust, seeing as they involved people putting things in their mouths that
shouldn’t be there, or their genitals into objects where they had no business being, or putting
objects of veneration in contact with objects associated with uncleanliness. Yuck makes wrong!
No wonder they found it hard to “translate.”

This phenomenon of dumbfounding was taken to indicate that the rationalist paradigm prevalent
at the time had it backwards. According to Haidt’s interpretation, the primary role of conscious
and deliberate moral reasoning is seldom that of a motive force that gives rise to moral judgments, but rather as a faculty that enters the picture after the fact in order to present intersubjectively convincing rationales for moral judgments that had already arisen on the basis of immediate, emotional reactions. Although Reason can still exert indirect causal influence on our moral judgments inasmuch as we are often loath to act in ways we feel unable to properly justify, the judgments themselves tend to be made prior to the apprehension of such considerations. And even this regulatory function can ultimately be grounded in emotion,

28 The dog had already died in a traffic accident, so it was in a way sensible not to let it go to waste. The chicken
had been bought fair and square at the supermarket, already slaughtered for the purpose of being eaten, and the
lewd acts were conducted in the privacy of the man’s own home. There were no witnesses to be offended by the
flag’s desecration, and the act had no adverse effect on the overall moral disposition of the perpetrator, etc.
inasmuch as social emotions like shame and embarrassment play a large part in motivating this desire for justifiability. On this perspective, Reason does indeed seem to be, if not the slave, then at least the legal consultant or PR representative of the Passions.

This is why we cannot build our line of argument on Haidt’s account alone. Much of his work, and related work by others, is relevant to us as grounding for the assertion that our case judgments are strongly affected by emotional salience, but they may be of limited use if we are to present the macro-intuitions as any different from the others. On Haidt’s model, the macro-intuitive value premises arise from our empathic faculties and aren’t in any way less emotional than yuck makes wrong or retribution, while same rules apply can be viewed as a product of the social emotions that make us disapprove of arbitrary and inconsistent judgments and react with anticipatory distress at the possibility that others should direct such disapproval at us. If we are going to argue with Singer that the macro-intuitions are “different from” meso-intuitions on account of not being emotional, we shall have to look elsewhere.

4.2.2 – The telltale amygdala

Whereas Haidt found our moral judgments to be emotion pretty much all the way down, Joshua Greene’s research points towards some of them being a lot more emotional than others. Back when we addressed the consistency criterion in 4.1.3, there was one glaring lacuna: trolleys weren’t mentioned at all. That omission was motivated by the desire to avoid overlap, since they’d have to be addressed here in any case.

Trolley cases – various hypothetical situations in which an agent has to decide whether a runaway trolley should kill five as opposed to one – have emerged as the standard format for thought experiments aimed at identifying the factors that sometimes make our intuitions overrule the ceteris paribus presumption in favor of minimizing overall harm. The first trolley case, originally due to Foot (1976) and variously referred to as Sidetrack or Switch, aims at testing the applicability of the intuition this text has taken to calling acts/omissions asymmetry. The hypothetical agent is in a situation where he can either permit five people – railway workers stuck in the path of an out-of-control trolley – to die by omission, or else save them at the cost of actively causing one death by diverting the trolley over to a sidetrack where it will run over a single worker who would otherwise have been safe. When this hypothetical case is presented, a solid majority clustered around the 80% mark judges it morally right to act so as to cause one death rather than omit to prevent five. So this would imply that whatever else we may believe about acts/omissions asymmetry, few find it plausible that the “exchange rate” between acts and
omissions should diverge by as much as a 1-to-5 ratio. Due to its accessibility and high initial plausibility, **acts/omissions asymmetry** is a natural early candidate when trying to account for the underlying rationales of many moral judgments, but it didn’t seem to have made much of an impact on people’s judgments in this case. Since we expect proposed moral premises to be **applicable** to our judgments as well as just initially plausible, these results count in the disfavor of **acts/omissions asymmetry** (as mentioned in 4.1.3, this is an intuition that doesn’t do all that well on the **consistency** criterion).

When people refer to the Trolley Problem, they generally mean the conjunction of **Sidetrack** with another case – due to Thomson (1985) – called **Footbridge** or sometimes **Fat Man**. In **Footbridge**, the parameters are similar to **Sidetrack** inasmuch as there is a trolley on the loose and the agent can either omit to save the five or act so as to save them at the expense of one. But in this case, the act of sacrificing one life for the sake of five doesn’t involve redirecting the trolley onto a sidetrack where a single workman has the misfortune of being located. Instead, it involves pushing a very overweight man right off a footbridge and onto the track below, where his considerable bulk will stop the trolley before it hits the workmen in its path. Compared to **Sidetrack**, the numbers now flip and an equally overwhelming majority judges **against** the permissibility of pushing the fat man for the sake of saving the five. If we had tried to test **acts/omissions asymmetry** by this second case alone, we might have thought it to be a highly **consistent** and applicable intuition, but – seeing as the same variable made no such impact on the verdict in **Sidetrack** – we know better. But the dramatic contrast between the cases tells us that there must be some highly significant factor that differs between them. And in this case, **double effect** appears to be a highly promising candidate. In **Sidetrack**, the death of the single workman played no **causally necessary** part in saving the five. It was necessary in the sense that we couldn’t keep the trolley from hitting the five unless we did something that would also lead to the one being hit. But if he could somehow escape from the sidetrack, we would be all the happier and the five would have been just as saved as if he remained in place to be run over. Not so with the fat man in **Footbridge**: if he should miraculously bounce right off the track and land somewhere safe, we would be back to square one as far as saving the five was concerned. But **double effect**, as it turns out, is not **consistently** applicable either. The **Loop** variation combines elements of **Sidetrack** and **Footbridge**. As in **Sidetrack**, one saves the five by diverting the trolley onto a sidetrack, but the sidetrack loops **back onto** the main track again, which in turn means that they would still have been run over **except** for the fact that there is a fat man on the looped sidetrack that will stop the trolley as long as it hits him. If **double effect** had been the crucial
variable, respondents would have judged the same way about *Loop* as in *Footbridge*, seeing as this rescue plan depended on the fat man being killed. But they don’t (Edmonds, 2014, pp.35-43). And then there is the *Trapdoor* variation. It is similar to *Footbridge* in nearly all respects except for the fact that the fat man doesn’t have to be physically pushed. Instead, the agent has to push a button – at distance – in order to activate a trapdoor in the footbridge, which will in turn relegate the fat man to the role of meat barricade with precisely as little regard for his rights or his end-in-himselfness as a forceful shove would. But respondents do not react to this variation as they do to *Footbridge*: although there are fewer who judge this act right than sidetrack (which does indicate that double effect has an influence), the propriety of sacrificing the fat man is still not rejected to anything near the overwhelming degree as in *Footbridge*.\(^{29}\)

At this point, many will have the same suspicions as Greene: people’s intuitions about these cases seemed awfully sensitive to whether or not there’s direct bodily contact involved. But if this is the crucial parameter, the corresponding intuition (hands off!, perhaps?) can certainly be called an example of the poorly “translatable” sort. What does seem plausible, however, is that bodily contact could have a great influence on how intensely negative emotions the case elicited. We are social beings with hardwired social instincts, and among those instincts are strong emotional dispositions not to take lightly on acts that harm others. But the more forceful among those dispositions aren’t really concerned with harm in the sense of a general concept, so much as with particular types of harm. Harms that are inflicted up close and personal are in a class of their own when it comes to sounding our emotional alarm bells. It’s not that we fail to rationally comprehend that it’s just as bad to be turned into a trolley-stopper by the push of a button as by a forceful shove from behind. It’s just that we can’t really bring ourselves to feel in accordance with this awareness.

Such was the hypothesis Greene set out to test: that the “characteristically deontological judgments” – that is to say, judgments to the tune of “even if it would have made sense from a harms/benefits point of view, it’s still profoundly wrong and impermissible!” – are driven by emotion rather than any properly rational process. That is to say: when we feel that we shouldn’t push the fat guy no matter what the cost-benefit analysis would imply, this feeling is precisely that – it’s a feeling, an emotion – rather than indicative of any profound or rationally compelling truths that our moral theories ought to take into account at the basic level.

\(^{29}\) More precisely, the numbers as reported by Greene have 87% of respondents condoning action in *switch* (double effect, not touching), 31% in *footbridge* (not double effect, touching), 63% in *trapdoor* (not double effect, not touching), and 81% in *loop* (not double effect, not touching) (Greene 2013, pp.215 and 222).
Greene’s earliest experiments involved the use of fMRI scanners to measure the respondents’ neural activation patterns, so as to examine how this correlated with the type of case presented and the type of response elicited. As it turned out, there were interesting correlations. If the cases were apt to evoke strong emotions (*Footbridge* as opposed to *Sidetrack*), this correlated with more distinctly emotional activation patterns and a greater likelihood of deontological-type responses. The most interesting findings, however, were the contrasting neural activation patterns between those who judged emotionally-evocative cases in a “deontological” way, and those who nonetheless judged them in a “consequentialist” way (i.e., push the fat man for the greater good). Although both respondent groups would initially display considerable activity in brain regions associated with affective processing, the “consequentialist” respondents sometimes showed a little less of it, and – more significantly – accompanied it with greater activation in brain regions closely associated with the conscious override of automatic impulses. This would imply that when respondents gave consequentialist-type verdicts about emotionally-evocative cases, those judgments did not arise from affective processing, but rather from a concerted effort to override them in favor of decision procedures that were consciously judged to be more apt. So these results are definitely relevant to the question of whether some intuitions are more “emotional” than others. As far as these cases were concerned, it seemed as if the consequentialist-type verdicts arose through a conscious effort to override immediate, more easy-flowing emotional responses (Greene 2013).

These early experiments were soon followed up – both by Greene and other researchers – with additional studies. There’s no room for delving deeply into this burgeoning field, but most studies seem to concur that the tendency to judge cases in a “deontological” way correlates with many factors that would indicate greater emotional salience. “Deontological” judgments tend to be made much faster than “consequentialist” ones, which is congruent with them arising from affective rather than conscious-deliberative mental processes. Respondents with a reduced emotional repertoire – patients afflicted with frontotemporal dementia, low-anxiety psychopathy or certain types of brain lesion – are considerably more likely to condone actions like pushing the fat man. That would certainly indicate that emotion plays a causal role. Test subjects have also displayed less “deontological” proclivities when given anti-anxiety drugs, and more when given medications known to increase anxiety (generally as a side effect, seeing as there wouldn’t be much of a market for a dedicated pro-anxiety drug). That would imply that relative intensity of aversive emotion may have played a part. The tendency to judge in a “deontological” way also correlated positively with indicators of physiological arousal (sweaty palms), and even how well-
developed the patient’s visual cortex was (this might have enabled the subject to imagine the situation more vividly). Significantly, the response time for “consequentialist” judgments in emotionally-evocative cases rose considerably when respondents were afflicted with cognitive load, whereas this factor made no difference for the speed with which “deontological” judgments were elicited. This would imply that the cognitive resources being monopolized played a part in giving rise to the former judgment class, whereas the latter arose from a process that had little need of them. (Greene 2013, pp.124-128)

Summing up, then, it seems that the judgments we view as characteristic for the deontological stance have a tendency to be elicited more quickly and through processes that aren’t affected by cognitive load, they are less likely to be made when affect is blunted or the context less vivid, and they tend to be accompanied by greater physiological arousal. These correlations may not be conclusive evidence, but they do seem sufficient to provide corroboration for our overall line of reasoning if we can demonstrate a plausible connection between them and the considerations from section 4.1.

4.2.3 – Interference

How do the two preceding subsections connect with our overall line of argument? Can they support the claim that macro-intuitions are different from meso-intuitions in the sense that the latter is more like the intuitions referred to by Greene and Haidt? As far as concerns Haidt’s interpretation, they might not be different. But when we want to know if they are different, we are mainly interested in whether they can be described as “different processes,” and Greene’s research seemed to indicate that this was very much the case. Even if Haidt should be right that the intuitions favored by utilitarians are also founded on emotion in a more distal sense, the process by which they are judged applicable to particular situations is one that seems considerably less connected with emotion than what holds true for the other intuition classes. When people judged in accordance with the utilitarian premises we call macro-intuitions, there was a tendency for emotional activation to either seem less evident (Sidetrack-type cases, certain respondent groups or test conditions) or else counterbalanced by conscious, cognitive effort. So if we are going to claim that macro-intuitions are “different from” Haidt/Greene-style intuitions in the sense of arising from “different processes,” it does

30 That is to say, the respondent is given some secondary task – remembering a number, for instance – that lays claim to part of his working memory and other capacities involved in conscious deliberation.

31 I.e. concern for the general well-being as a kind of empathy-writ-large, concern for consistency as based on various social emotions (indignation at inconsistency in others, embarrassment at being subjected to the same by others), and so on.
seem as if Greene’s research can provide a good grounding. But do we have a strong enough grounding for equating the meso-intuitions with the ones studied by Haidt and Greene?

As we recall from Chapter 3, the meso-intuitions were defined by their abstract/concrete discrepancy, which we interpreted as a symptom of them “borrowing” intuitiveness from the contexts they were associated with. We didn’t specify what we meant with “concrete intuitiveness,” but emotional salience seems a plausible candidate.\(^\text{32}\) A lot of meso-intuitions do seem to overlap or relate closely with emotions. Chapter 2 described our concrete and abstract intuitions as arising from “different processes with different functions.” Our concrete intuitions were tasked with “practical guidance within familiar parameters,” whereas our abstract intuitions were subject to fewer restrictions of that kind. What could this be taken to imply in a more specific sense? Greene’s work may hold a few leads.

The previous subsection focused on the correlation between emotional salience and the likelihood of “deontological” judgments, without going into detail about the mechanisms involved beneath the cranium. But this is quite an interesting matter, and may be worth a closer look now that we are concerned with “different processes.”

Although it goes without saying that this is a simplified outline, there were four brain components that seemed to display especially prominent activity when test respondents adjudicated high-emotion cases like Footbridge: the amygdala, the ventromedial prefrontal cortex (VMPC), the anterior cingulate cortex (ACC), and the dorsolateral prefrontal cortex (DLPC). All of the non-pathological respondents displayed considerable activation in the two first components right on the outset: the amygdala (which plays a central role in giving rise to emotions – and especially aversive ones like disgust and fear) and the VMPC (central to processing emotions). As for the ones who eventually did judge high-emotion cases in a consequentialist way (e.g., pushing the fat man), they followed up this initial activation of the “emotional” parts with significant activation of the other two components – the DLPC and the ACC.

Evolutionarily speaking, the DLPC is our most recently developed brain region, and the one most intimately connected with conscious reasoning. Among the myriad mental processes

---

\(^\text{32}\) Or at least one of them, seeing as there’s no need to commit to the position that it should only be emotion involved and nothing else.
that involve the DLPC,\textsuperscript{33} there is one class that holds particular significance to this context – the
processes involved in \textit{cognitive control}. By cognitive control, we mean the capacity to
coordinate or \textit{orchestrate} different mental processes. That is to say: we exhibit cognitive control
when we consciously ensure that the \textit{appropriate cognitive process} is applied to the task at hand.
This can be illustrated more clearly by a classical test used to measure it: \textit{the Stroop task}.
Various words are exhibited to the respondent, written in different colors. The respondent’s
instruction is to name the color of the word being exhibited. However, this task is made far more
challenging by the words themselves being \textit{names of colors} that are \textit{different} from than the ones
they are written in. Thus, the word “blue” is displayed in red letters, or the word “red” in green
letters, and so on. It takes a lot of concentration to get this right, because the process of reading a
word aloud is something that comes with far more fluidity and ease than the process of observing
a word, noting the color of the letters, recalling the name of the color and then stating the name
of the color aloud. As for the other, not-so-automatic process of naming the color, it would itself
be a very easy task under \textit{normal} conditions, but becomes very difficult to perform when
subjected to interference from the more-automatic process. In other words, the Stroop task puts
the respondent in a situation where a task that would normally be very easy (name the color of
the letters) becomes very challenging because it is now being attempted in a context that triggers
a \textit{competing, more automatic} process that is not apt for the task consciously pursued. In order to
avoid acting along with the automatic response triggered by the context, the respondent must
exhibit \textit{cognitive control}: he must be capable of recognizing certain more-automatic processes as
inapt for the task at hand, and overrule them in favor of other, less-automatic processes that are
apt. And here’s the clincher: when respondents to the Stroop test are having their brain activity
monitored with fMRI, the observed patterns display remarkable similarities with those of
Greene’s respondents who made “utilitarian” judgments about emotionally-evocative cases
(Greene 2013, pp.119-121).

Greene’s theory about these experiments is, put simply, that the underlying processes are
quite analogous to what’s going on in the Stroop task (Greene 2013, pp.120-121). To judge that
it’s appropriate to minimize deaths in \textit{Sidetrack} is like naming the color of a word that is \textit{not} a
color word: there is no strong competing impulse triggered, so the respondent readily goes with
the decision procedure that seems appropriate on conscious reflection. In cases like these –
where the context doesn’t trigger very strong emotions – nearly everyone accepts the

\textsuperscript{33} And the ACC as well, but we’ll make it simple by just focusing on the DLPC. For the record: the ACC is
known to be implicated in a lot of error-detection functions, and it generally works in concert with the DLPC
when performing those.
consequentialist rationale. More so – and this also relates to the criteria evaluated in 4.1 – none of Greene’s respondents were confused or unclear about the underlying, consequentialist rationale of minimizing overall harm. Those who judged in accordance with it never presented any other rationale, and those who judged that it ought to be overruled still recognized it as relevant. In contrast, respondents often gave widely varying rationales for judging high-emotion cases in a “deontological” way, and sometimes rationales that were outright confused\(^{34}\) (Greene 2013, p.128). We are in familiar territory here, seeing as 4.1 has already treated the criteria of universality and related issues in a fair bit of detail. The fact that the consequentialist type of rationale is recognized universally as plausibly relevant and accounted for with greater clarity and consistency would imply that it’s a type of decision procedure that is consciously recognized as apt, similar to how the decision procedure of “observe color-recall name of color-state name of color” is consciously recognized as apt in the Stroop task. On this analogy, a respondent presented with a high-emotion case – like Footbridge – is a bit like a Stroop task respondent who has just been presented with the word “blue” written in purple letters. In the latter case, the use of the color word “blue” triggers a near-automatic response that interferes with the otherwise-easy task of describing the color of the letters. In the former case, the vividly unpleasant emotions engendered by the thought of forcefully shoving someone off a footbridge triggers an automatic response that interferes with the otherwise-easy task of applying the decision procedure “do what will minimize harm.”

If this account should be correct, it could be taken to imply that utilitarian-style moral reasoning is the “default” type of moral reasoning – as indicated by its greater universality, abstract plausibility, consistency and precision/clarity – whereas deontology-style moral reasoning is the kind that results when the particularities of certain contexts trigger strong emotions that interfere with the judge’s ability to apply such standard decision procedures. Hardly an uncontroversial claim, but – it does seem to cohere pretty well with the various considerations we’ve been going through in this chapter. The criteria in 4.1, the tendency for people to get “dumbfounded” when they cannot connect their judgments with macro-intuitive rationales, and finally – as we saw here – the conspicuously Stroop-like neural signatures

\(^{34}\) For instance, there was a recurring tendency for respondents to justify their deontological verdict in Footbridge in terms of acts/omissions asymmetry rather than double effect. One may suspect that there is a slight Stroop-like dynamic to this tendency as well, seeing as respondents are likely to have intuited that causal structure had something to do with their judgment, which may then have led them to invoke the more easy-flowing rationale of the acts/omissions distinction (also about causal structure so liable to be triggered by implicit association, but not applicable to the case) rather than the less easy-flowing intending/foreseeing distinction (about causal structure, legitimately applicable to the case, but more inaccessible or difficult to explicate for non-philosophers).
displayed by those who restrained themselves from sliding into competing, more easy-flowing forms of cognition that might in turn be found hard to “translate.” The pieces of this puzzle do seem to connect.

4.2.4 – Different processes with different functions

The aim of section 4.2 is to obviate the presumed baseline credibility of the meso-intuitions. So we have to look a bit closer into why their connection with emotion should give us reason to view them as less trustworthy. It’s not obvious that this should be the case. People with a diminished emotional repertoire – like the patient groups recruited to some of the trolley studies described in 4.2.2 – aren’t at all doing very well as far as moral judgment is concerned (Woodward and Allman 2007). To presume that the more easy-flowing emotional response is as inapt for the decision task of Footbridge as the more easy-flowing read-the-word response for the Stroop task is to beg the question. Whether or not it’s morally appropriate to push the fat man is an open question, and the fact that those who judge it inappropriate seem to do so on emotional grounds is only undermining to the degree that we have reason to think the emotions in question are inapt. And it’s not at all obvious that it should be inapt to react with horror at the thought of pushing someone off a footbridge or disgust at the thought of carnal congress with a dead animal. Clearly, we have some justificatory work to do here.

We have to take a closer look at those brain regions mentioned in 4.2.3. The tendency to judge in one way as opposed to another seemed to correlate with the involvement of different brain regions, some of which were phylogenetically “younger” than the others. The DLPC – the brain region so heavily involved in the Stroop task – is the youngest part of our brain, whereas the more deontologically-inclined VMPC is somewhat older (but still pretty young as these things go), and the amygdala positively ancient in comparison. Where cognitive processes are concerned, evolutionary age correlates positively with efficiency and negatively with flexibility (Greene 2013, pp.132-143). Other things being equal, evolution will favor action-guiding mechanisms that waste as little time and energy as possible. Thus, the instincts and reflexes that emerge from the oldest parts of our brains – the brainstem and cerebellum and those other regions that are so old that we share them even with the reptiles – are extremely quick and entirely automatic. But whereas the reptiles have generally stuck with the efficient-but-inflexible strategy, the mammalian tendency has been towards greater flexibility and adaptability even at the cost of “hungrier” brains and occasional hesitation. Hence, we have emotion as a characteristically mammalian form of mental processing, and the limbic system as a characteristically mammalian brain structure. Emotions aren’t as quick as instincts, but they are a
good deal more flexible and context-sensitive. The emotional processing of humans is an extremely sophisticated epistemic resource, but even that capacity couldn’t fully provide the flexibility needed to face the adaptive challenges of our hominin ancestors. Thus, the capacity for conscious reasoning emerged: extremely slow and arising from brain components that gobble up a solid 20% of our total caloric intake, but with an unprecedented capability for adapting to novel situations, and capable of such neat functions as planning, comparing alternatives, and restraining ill-suited impulses (Passer and Smith 2008, pp.174-175; Cartwright 2008, p.119). So as long as we value that kind of flexibility more than we value efficiency, newer will be better.

But it cannot be taken for granted that flexibility will always take precedence. When we speak of efficiency, we aren’t just concerned with speed but also with the capacity to take many variables into account. When making practical moral judgments in real-world contexts, it is crucial that we are aware of as many of the relevant parameters as possible — likely consequences, the motives and intentions of others, the emotional impact our acts will have both on observers and ourselves, the prevailing norms around us (including unspoken ones), and so on. And it’s not just that these factors are many, but also that they interrelate with each other in complex ways, and are unpredictable enough that one ought to take several possible outcomes into consideration. If someone were to attempt a proper and adequate assessment of all these factors with conscious reasoning alone, the process would be extremely time-consuming, and it’s quite likely that he would end up overlooking some elements anyway. Unlike reasoning, however, emotions aren’t limited by the capacity of our working memory. By using labor-saving heuristics and bypassing the involvement of conscious deliberation, our emotions can make a far more satisfactory assessment of these factors. Thus, it’s not surprising that the patient groups with diminished emotional repertoires are notoriously bad at moral judgment in real-world situations: those situations involve far too many variables to be adequately analyzed with the limited processing capacity of Reason alone. So our emotions are indeed well suited for these functions (Woodward and Allman, 2007, pp.179-202).

Those functions, however, are different from the ones involved when we deliberate about morality in the abstract, in the context of formal theory. At this level of analysis, there are different things we want to achieve, and we may find that the labor-saving shortcuts that make emotions indispensable for practical moral judgment now turn into defects. Our emotions owe much of their speed and capacity to the fact that they don’t bother to keep our conscious mind informed of how they arrive at their verdicts, and neither do they worry about things like internal consistency, generalizability, precision, or verbalizability. So it’s not just that it’s hard to
“translate” emotionally based moral judgments, but also that there’s no particular reason to expect them to be translatable at all.

And although emotions are very sensitive to contextual factors, they may not be very sensitive to the difference between context-specific as opposed to more general relevance. When we do moral theory, we want to distinguish between those factors that are contingently relevant in a certain context due to other, underlying considerations, and those factors that are relevant on a wider and more basic level. But it may be a little optimistic to think our emotions should be well-suited for making this distinction. In particular, it’s hardly obvious that our emotions should be capable of disregarding our usual background assumptions whenever we hear the phrase “suppose for the sake of argument.” Our conscious reasoning may apprehend that under the very unusual circumstances of Footbridge, it’s hard to give a good account of why our reasons for pulling the switch in Sidetrack shouldn’t be equally valid as reasons for pushing the fat man. Not so with our emotions: their purpose is practical guidance under familiar parameters, and as such they forcefully and immediately tell us that pushing the fat man is out of the question. And about this, our emotions are in a sense right. But what’s less certain is whether our emotions are being right in a way that tells us anything theoretically useful. They are right in the sense that one shouldn’t push the fat man in real life, and they are right in the sense that they exemplify a disposition that it’s better on the whole to have than not to have. But those aren’t the things we try to find out when we do trolleyology. We had hoped to discover some underlying, less obvious principle that might explain our strong intuitions about the case even if we should postulate away all the contingent, real-world reasons for why one shouldn’t push the fat man. And our emotions don’t play that game very well: they are fast-and-loose heuristics that tend to presume ecological validity, and they will not suddenly turn off that presumption whenever someone says “suppose we know for certain that the outcome will be so-and-so.” So in thought experiments with highly unusual postulates, it is quite likely that our emotions will refuse to play along and rather react as if the parameters of the situation were more realistic.

And then there’s modular myopia. This is a term Greene uses when describing his theory about the underlying mechanisms of double effect, although its applicability certainly extends to other intuitions as well. In evolution-speak, a module is a cognitive subsystem tasked with some fairly domain-specific function, so it’s a bit like the mental equivalent of a smartphone app. Significantly for our purposes, modules also tend to keep our conscious minds in the dark about their inner workings, which makes them prime candidates for giving rise to the kind of non-inferential judgments we call intuitions. Greene’s theory holds that our minds include an “action-
plan inspector” module, tasked with keeping an eye on our planned behavior and protesting strongly against action plans that involve casual violence. On this account, it is this module that’s been intervening when we feel so awful about the idea of pushing the fat man, as well as a great many other prototypically violent actions that it’s adaptive to abstain from. But the action-plan inspector module doesn’t register and object to every violent act we contemplate – it is myopic, after all, and cuts too many corners to notice everything (harm caused by flipping a switch is likely to fly below its radar, for instance). However, another factor that can let violent acts evade detection by this module is when they branch off from the primary action plan. That is to say: the module inspects the primary sequence of intention-means-result and will sound the alarm if there is violence involved here. But if the violence in question is a side effect of the means, then it essentially branches off (on a sidetrack, one may be tempted to say) from the primary action plan and is outside of the module’s detection range. Our conscious reasoning is still capable of understanding that we are intentionally harming someone, but the module doesn’t get it, which means that it’s not going to sound the emotional alarm bells that it otherwise would. So even if we rationally comprehend that it’s just as bad to be run over by a trolley on a sidetrack as on the main track, it just doesn’t feel as impermissible to subject someone to the former as the latter (Greene, 2013, pp.224-240).

A related problem is the tendency for emotions to rely on the prototypicality heuristic. That is to say, they tend to save time and effort by treating certain stimuli as “blanket triggers” for responses that are often but not always appropriate. Garden hoses are prototypically snakelike in a way that will sometimes trigger startle reflexes: if our brains notice something that might be a snake, they are liable to startle first and double-check afterwards. This is a pretty adaptive response, seeing as it’s more dangerous to mistake a snake for a garden hose than a garden hose for a snake. Nonetheless, our brains are still not getting it right here, and we wouldn’t take our reactions in these cases to be telling us anything useful about herpetology. Likewise, an instinctive reluctance to harm people up close and personal is adaptive, but it needn’t follow from this that an adequate moral theory ought to treat it as indicative of some fundamental principle that needs to be accounted for. Quite to the contrary, it seems more apt for a moral theory to take particular care in reminding us that we shouldn’t take any more lightly on distantly inflicted harms. It is when we cannot reliably trust our moral instincts for guidance that we need theory the most, after all.

There are a lot of other issues that could have been addressed as well, but we’ll have to proceed to the next section. Suffice to say that there are a great many ways in which emotional
processing achieves its “first-person” efficiency at the expense of the kind of decontextualized, generalizing perspective most conducive to the criteria we examined in 4.1. Affective processing is a splendid epistemic resource that often serves us well in our daily lives, but it’s doubtful that it can serve a very satisfactory role as a supplier of basic premises in formal moral theory. We are now at a point where we may outline our second rebuttal to the Only Game in Town objection.

**Rebuttal #2: Wrong processes for the wrong functions.** Unlike macro-intuitions, meso-intuitions derive part of their plausibility from the emotional salience of the particular cases they are associated with. This is problematic, because emotions are not well-suited for the functions that we would like a moral datum to serve in the context of formal theory.

It is now time to proceed to the final section and wrap up.

**Section 4.3 – Conclusions and implications**

**4.3.0 – Did we achieve our aims?**

Now that we are nearing the end, we shall recall the beginning by revisiting the three aims stated in 1.0.1. They were as follows:

a.) To make a *general* case for why and how we ought to limit the reliance on intuitions.

b.) To connect this *general* case with the more *specific* debate surrounding the justifiability of utilitarian intuition-deployment practices.

c.) To contribute to greater terminological precision.

The purpose of this subsection is to see if the preceding text has brought us any closer to achieving them.

**Chapter 2** focused on the “why” side of the *general* case, drawing on history to demonstrate the pitfalls of letting intuitions serve as an alternative to explicit rationales for the moral distinctions we make. Time and again, it has been concluded that there really *was* no justification for the distinctions then considered natural and proper. There really *was* no morally valid or significant nuance being ignored when Mill saw no distinction between the civic entitlements due to women and men, Bentham’s theory left no opening for letting one’s own pleasures take precedence over the presumably coarser pleasures pursued on the continent next door, and the Commission failed to distinguish between the disapprobated but nonpunishable...
conduct of the adulterer and the disapprobated and punishable conduct of the homosexual. The simple phrase “why should that make a difference?” has always been a bugbear to unjust and arbitrary practices, serving as a reminder that it is the one who lets it make a difference who owes an explanation. But against this bugbear, the hoary injustices of the past kept taking refuge in the bastion of our intuitions, which were ever eager to confirm that there were no further reasons needed to continue keeping apples and oranges apart. Between our intuitions’ track record (2.1), affinity for the “apples and oranges” perspective (2.2), and their potential to serve as alternatives to rationales more amenable to discussion and revision (2.3), there seemed to be good reasons to prefer not ascribing them any greater evidentiary role than necessary.

Thus, we set out to defend the intuition-minimalist position against the objection we called “The Only Game in Town.” The term “intuition” is broad, and it can equally well be applied to the abstractly-compelling notion that the onus probandi falls upon the one who would not let the same rules apply as to the more contextually-salient tendencies to judge certain cases as clearly incomparable even if we have a hard time explaining why. If we are to be justified in distrusting the latter type while still relying on the former, it is necessary to outline how they are relevantly different. Are we now in a position to do this?

I would say yes. Directed by our second working hypothesis – the assumption that we can learn something useful from the intuition-deployment practices of classical utilitarianism – we set out to examine and classify a broad range of moral intuitions throughout Chapter 3, thereby pursuing our third aim of terminological expansion as well. We landed on the abstract/concrete discrepancy as a promising candidate for what made the intuitions in 3.1 different from those in 3.2 in a way that could justify privileging the former class. We then proceeded to see if this candidate would deliver on its promise as Chapter 4 set out to compare the methodological merits of the intuition classes that did and did not display abstract/concrete discrepancy – the meso- and macro-intuitions respectively – in accordance with the criteria presented in connection with our first working hypothesis in 1.2.4. After finding that the meso-intuitions all seemed to have some problem or other along the interconnected dimensions of universality, precision/clarity, consistency and abstract plausibility that made them less ideal for theoretical purposes than the macro-intuitions, we outlined our first, fairly pragmatic-methodological rebuttal to the Only Game in Town, dubbing it R1: The opposition is split in recognition of how the meso-intuitions only seemed to present a united front as long as they remained vague. But as this argument seemed a bit thin on its own without a corollary account of what underlying problem these pragmatic-methodological defects could be indicative of, we proceeded to
investigate this in 4.2. Taking as our lead the Singer quote from 1.2.3, we asked what it could mean to call a type of intuitions similar to the ones referred to by Greene and Haidt. First we looked at the intuitions studied by Haidt, which were tightly interconnected – perhaps even identical? – with emotion. Upon finding that the macro-intuitions wouldn’t really be any “different from” the meso-intuitions as far as Haidt’s interpretation was concerned, we continued by looking at Joshua Greene’s work. This was more promising, as it seemed to indicate that the underlying processes involved were very different. Although it gave some slight pause that emotions seem very important to practical moral reasoning, we soon concluded that the properties that made them well-suited for this function did not make them equally well-suited for formal moral theorizing. We were then ready to outline our second rebuttal, which we called R2: Wrong processes for the wrong functions.

We are now ready to combine the two rebuttals outlined previously in the chapter into a third and final compound rebuttal against the Only Game in Town objection.

Rebuttal #3 (Final): The opposition is split because it uses the wrong processes for the wrong functions. The utilitarian distrust of moral intuitions as a general class is not inconsistent with their reliance on certain, innately-cogent premises that are also classifiable as intuitions. This is because the intuitions in the latter class are macro-intuitions, and as such possess qualities that make them different from the other classes in ways that are epistemically and methodologically relevant. Compared with other classes of intuition, they display considerably greater ceteris paribus consistency of applicability, abstract plausibility and universality of acceptance. This is not to say that all the implications derived from them possess these qualities, but when such implications are found inconsistent, implausible or contentious, this is generally regarded as the result of some countervailing consideration rather than as disproof against the macro-intuitions themselves possessing these desirable methodological qualities. More so, the macro-intuitions do not depend on being accounted for in vague terms in order to retain their plausibility, consistency or universality, but will be found so also when outlined in clear and precise terms. The classes of intuition that utilitarians distrust do not possess these methodologically desirable qualities to the same extent. The intuition classes they disregard include micro-intuitions and meso-intuitions. Micro-intuitions are singular and not viewed as freestanding warrant for premises anyway, and are known to be subject to many possible distorting factors (Sinnott-Armstrong 2011, pp.15-17). As for the class of meso-

35 If they are generalized or in other ways supply the basis for a more general premise or notion, that would count as a meso-intuition (provided it’s at all intuitive in the abstract).
intuitions, these are characterized by their failure to seem equally plausible in the abstract as when elicited in particular contexts. This may be taken to indicate that part of their plausibility derives from factors that are specific to the context, which would compromise their suitability as warrant for methodologically basic premises. Many empirical studies indicate that the emotional salience of the particular contexts meso-intuitions are associated with plays a significant role in compensating for their shortfall in plausibility when considered purely as abstract notions. Although meso-intuitions are not uniform in how they fare according to the criteria of universality, consistency and abstract plausibility, there is a notable tendency for most intuitions of this class to depend on being vague in order to retain such qualities fully. Trying to account for them in clear and precise terms will often diminish the other, methodologically valuable qualities. These problems seem congruent with the suggestion that they are relying on emotion for part of their plausibility. Although it need not follow from these differences that the other intuition-classes ought to be discounted as worthless, their relevance is significant enough that it cannot be called arbitrary or unmotivated to treat the macro-intuitions as epistemically or methodologically privileged.

Mission accomplished? I shall say myself satisfied, at any rate.

My third and final aim was to contribute to greater precision and clarity in the vocabulary around intuitions. This aim has taken center stage in the descriptively-focused Chapter 3 and borne fruit through the outlining of the tripartite taxonomy.

4.3.1 – Further implications

How’s for our working hypotheses? The first one held that it would be wise if we distinguished more clearly between descriptive and normative success criteria in ethics. There was a certain paradoxical tension between this claim and the criteria I listed up in 1.2.4 and eventually applied to the intuition classes in 4.1, for these criteria – precision/clarity, universality, consistency, and abstract plausibility – could all be described as descriptively relevant success criteria as well. Nonetheless, I did set out with the assumption that some of these criteria might prove more valuable for strictly normative pursuits than others. The only truly normative criterion was the one I called the Seventh Criterion, and that was an emergent criterion that could presumably be achieved through a correct balancing of the others. My second working hypothesis held that the historical track record of classical utilitarianism seemed indicative of its having hit upon such a correct balance, which in turn led me to examine what made the intuitions relied on by this system different from the others. As it turned out, it may not
so much have been a matter of balancing as it was a matter of only basing one’s premises on the few intuitions that seemed to satisfy all four of criteria 3-6. This rather sparse list of foundational axioms led to a system that had few safeguards against the sort of implications that could put its descriptive fidelity in jeopardy, but simultaneously served well in restraining the tendency towards undergeneralization so often found to have been obstructive to the pursuit of the Seventh Criterion. As for the other intuitions, they all seemed to satisfy some of the criteria but never all, apparently making up for the shortfall by borrowing plausibility from the particular contexts they were applicable to. Empirical data seemed to indicate that emotional salience might have supplied a large proportion of this borrowed plausibility, which seemed highly congruent with the tendency for their other desirable methodological qualities to slip away if one tried to make them clearer and more precise. This may imply that the assessment of an intuition’s precision/clarity could be a key strategy for getting an inkling of how much its initial compellingsness results from emotional salience as opposed to a more “proper” form of self-evidence.

These considerations – abstract/concrete discrepancy and a failure to be simultaneously clear and certain – could be worth keeping in mind when outlining the requirements for an intuitive judgment’s status as “considered” enough for inclusion among the initial input of Reflective Equilibrium. If an intuition is such that it never quite seems satisfactory unless cloaked in vagueness or elicited in a context, that may be a good indicator that it really isn’t very satisfactory at all.

Before we conclude, it’s worth addressing a final element. Are the arguments throughout the preceding text meant to imply that the meso-intuitions ought to go? Most of them do seem pretty reasonable, after all, and not necessarily as if there’s much harm in complying with them. In many cases, that may be true. But just as with the emotions many of them may be relying on for phenomenological fire support, these intuitions may not be so well suited for theoretical purposes. If we view our theories as under obligation to accommodate intuitions that really aren’t responsive to any rhyme and reason, we’ll end up chasing our tails in perpetuity. And some of these intuitions can have a pretty negative influence. Acts/omissions asymmetry and loss-no gain asymmetry can have a pretty conservative slant, for instance. It would undoubtedly do an immense amount of good if people had been more willing to judge a politician’s failure to reform outmoded laws as harshly as they would if he had actively introduced them, and the fact that we don’t may be a fact about our epistemic limitations that our formal theories would do better to correct for than enshrine. When politicians and generals opt to permit the bombing of
civilians, they always describe their intentions in terms of \textit{double effects} anyway (that’s why they call it “strategic assets denied to the enemy” rather than a leveled city block), so it’s unclear how much of a \textit{constraining} role this principle serves at all. But what it may be doing is to act as a moral fig leaf – an unacknowledged prerogative – for carrying out policies that we would clearly recognize as unjustified if we had evaluated them on the basis of consequences alone. If a lawmaker suggested that we ought to deter criminals by threatening their families, we would immediately recognize this as inappropriate. But as long as the \textit{direct} means of deterrence is imprisonment of the criminal while the suffering this will cause to his family is viewed as a merely foreseen side effect, it becomes tolerable all of a sudden. And then there’s \textit{charity begins at home} – if we fail to be moved by the plight of starving children on other continents while simultaneously recognizing their unconditional claim to help \textit{if only} they had been drowning in our immediate vicinity, then it’s no reason to be surprised the Mediterranean claims ever more lives by the year. This is downtown “apples and oranges” thinking. I’m all for keeping the fat man safe and all that, but he is a red herring inasmuch as \textit{Footbridge} is one of those cases where our emotions function the way they’re supposed to. It’s in the kind of \textit{Sidetrack}–type, flying-beneath-the-radar cases like the ones above that we \textit{really} need moral principles.

This concludes the text.
References


