Chapter 2
Metaethical Foundations

In the previous chapter, we narrowed down what we mean by the term "moral theory," and we developed an understanding (at least a preliminary one) of what it means for a moral theory to be consequentialist. Since it is our goal to criticize all consequentialist theories, we should, in a next step, address the question how we can evaluate them. This is what we shall do in this chapter.

In Sect. 2.1, we will introduce an influential approach to theory evaluation which we will refer to as the Rawlsian Approach. It can be factorized into at least three evaluative criteria, viz. consistency, connectedness, and intuitive fit. On the Rawlsian Approach, we can criticize moral theories by pointing out that they leave something to be desired in regards to at least one of these criteria. Since the primary objections to consequentialism draw on intuitive fit, we shall focus on this sub-criterion alone.

We can distinguish between three interpretations of the criterion of intuitive fit, viz. the Top-Down Approach (TD), the Reflective-Equilibrium Approach (RE), and the Bottom-Up Approach (BU). In Sect. 2.2, we will discuss these three approaches. It will be our aim to establish that we can reject TD and that either RE or BU is justified. This will play a crucial role in our argument. Here is why. To refute consequentialism, we will draw on our moral intuitions about individual cases. Such a procedure is admissible both on RE and on BU, but would be ruled out by TD.

Having clarified the interpretation(s) of intuitive fit on which our argument relies, we will proceed to develop this criterion into a workable method for our investigation. This is necessary for the following reason. Intuitive fit merely states a philosophical ideal, viz. that our moral theories should fit our intuitions. It does not, however, give us a methodic procedure that we can use to test whether a given moral theory does, in fact, live up to this ideal. In Sect. 2.3, we will, hence, discuss how we can apply intuitive fit in moral argumentation. Our answer to this question is the Provisional Fixed Point Approach (PFPA). On PFPA, we evaluate moral theories as follows. We look for very strong intuitions about cases – provisional

N. Mukerji, *The Case Against Consequentialism Reconsidered*,
Theory and Decision Library A: 51, DOI 10.1007/978-3-319-39249-3_2

fixed points – and examine whether a given moral theory can match them. If not, we reject it (subject to the *proviso* that the best moral theory is, in fact, compatible with these provisional fixed points).

PFPA leaves open which kinds of cases we should use. In Sect. 2.4, we will introduce 'trolley cases' which we will use in our argument against consequentialism. We will discuss their characteristics and possible uses. Since there have been many objections to their applications in moral philosophy, we will also discuss their *pro*s and *con*s.

In Sect. 2.5, we will close the chapter with a brief summary of the main points.

2.1 The Rawlsian Approach

How can moral theories be evaluated? When we ask this question, we leave the field of normative ethics and set foot into the area of metaethics and moral epistemology, in particular. This is worth emphasizing. After all, at the beginning of the second chapter we characterized the subject of our inquiry as a matter of *normative* ethics. Our digression, however, seems justified.[1] Our goal is to develop a convincing critique of consequentialism. To do this, we need evaluative criteria. After all, every objection to a moral doctrine is a claim that it falls short of a particular evaluative criterion.

Now, which criteria should we use to evaluate moral theories? One obvious answer is to say that a theory (or, at least, its theoretical component) should be *true* and that we should, hence, adopt *truth* as our evaluative standard. This approach, however, is not terribly fertile. Moral philosophers are very much divided on the issue of whether or not moral theories can be true. *Moral realists* affirm this, while *moral anti-realists* deny it.[2] However, even if there was agreement on the matter of moral truth, it appears that this would not help us much. For the issue of the truth of moral theories – if, in fact, it can be had – might be largely independent of the question whether we should accept them (cf. Railton 1984, 155). To see this, consider the analogous controversy between scientific realists and instrumentalists in the philosophy of science. A stylized picture of their debate looks like this: *Scientific realists* believe that our scientific theories ought to convey the truth about the world, presupposing, of course, that these theories

[1] On this point, see also Pettit (1997/2007).

[2] An instructive discussion of the realist position is offered by Sayre-McCord (2011). For a concise general examination of anti-realism, see Joyce (2009). Moral anti-realists are commonly partitioned into non-cognitivists and error theorists. Non-cognitivists believe that our moral persuasions are not apt for truth or falsity. They suggest that they are, rather, expressions of emotional attitudes (cf., e.g., Barnes 1934, Ayer 1952, 102–120 and Stevenson 1937) or prescriptions (cf., e.g., Hare 1961; Gibbard 1990). Error theorists believe that moral views can have a truth value, but think that all our moral judgements are false. The classic statement of such a view is found in Mackie (1977).

can have a truth value. *Instrumentalists*, on the other hand, deny that scientific theories are apt for truth. They believe that they do not refer to something *real*, but are rather devices for predicting observable phenomena. It seems, then, that scientific realists and instrumentalists are at an impasse when it comes to the issue of theory evaluation. Realists will evaluate theories regarding whether or not they are true, while instrumentalists will assess them in terms of their predictive power. However, this picture is not accurate. It is important to keep apart the issue of theory *acceptance* and matters of *ontological interpretation*. Both realists and instrumentalists, it seems, can *accept* theories based on the same criteria since "[t]he acceptance of a theory involves only the claim that it is empirically adequate, not its truth on the theoretical level." (Niiniluoto 2011) Nida-Rümelin (2002, 45) emphasizes this point, too, and throws a bridge to moral theory. For reasons of space, we shall not go into the reasoning he gives. Rather, we shall simply take it for granted that the distinction between the epistemological issue of theory acceptance and the ontological problem of a theory's aptness for truth, as drawn in the philosophy of science, carries straight over to moral philosophy. In our inquiry, then, we shall put aside questions about moral truth and turn immediately to the evaluative criteria for moral doctrines.

Alas, the field of moral epistemology, which deals with these criteria, is also highly controversial. Hence, any stipulations we might make about the criteria of evaluation for moral theories are bound to be controversial as well. Unfortunately, though, we have to make at least some such stipulations. For, plainly, "[i]f we take up a point of view stripped of all evaluative conviction, we have no basis for evaluation." (Hooker 2003, 11).

There are various contrary viewpoints about the justification of a moral theory. Some theorists suggest that it is justified if it conforms to the *will of God* (e.g. Quinn 1990). Accordingly, the method of evaluation might consist in comparing the content of a moral doctrine with the laws that are laid down in some sacred text. Alternatively, it may consist in personal revelation. Moral naturalists maintain that ethics should be grounded in *empirical facts*. These theoreticians may favour a scientific study as a way of making progress on moral questions.[3] Others, most notably Kant (1785), advocate a rational approach to ethics which regards *pure practical reason* as the ultimate arbitrator on matters of right and wrong. A further approach to assessing moral doctrines is the *intuitionist method*. It assumes that we can know certain moral 'facts' simply by intuiting them (e.g. Ross 1930/2002; Prichard 2002; Crisp 2006). Another view on moral justification is due to Hare (1981). He believes that we can support his moral theory (a version of preference utilitarianism) through a careful analysis of the meaning of moral language and, in particular, the property of universalizability that, as Hare argues, attaches to moral utterances.[4] Nowadays, however, the most common conceptions of moral justification appear to be variants of what may be called the Rawlsian Approach.

[3]For a summary article on moral naturalism, see Lenman (2008).

[4]I am grateful to Julian Nida-Rümelin for pointing out to me that my argument does not cover versions of consequentialism that follow Hare's justificatory approach.

Like the intuitionist method, it also allows intuition to play an important role (cf. Rawls 1951, 1971/1999). Due to its status as the standard view of justification in modern ethics, we shall adopt it. Given the task that lies ahead of us, we shall not, however, attempt to justify it. The remainder of this section shall serve merely to explain and interpret the Rawlsian Approach.[5]

In his famous book *A Theory of Justice* (1971/1999), John Rawls starts his discussion of the issue of justification in moral theory by pointing out that human beings have a remarkable faculty. In the context of his theory of justice he calls it a "sense of justice." More generally, we could call it a "moral sense." Now, this moral sense can be thought of as the capacity to form moral intuitions at various levels of generality. We are capable of having high-level intuitions about abstract moral principles that cover a whole range of moral cases or even all cases. Moreover, we can form low-level intuitions which cover only a narrow variety of scenarios or, in the limiting case, just one particular moral problem.

We may regard our moral sense as analogous to our sense of vision. When we consider a particular moral judgement we can *sense*, as it were, whether it is correct. In a similar way, one may say, we can *see* whether an object has a particular colour (cf. Harrison 1967, 72).[6] This analogy brings out an important point about moral intuition. Sometimes we choose not to believe what we see. Similarly, we may sometimes opt to disbelieve our moral intuitions (cf. Greene 2008, 63 and Sinnott-Armstrong 2008, 48). To make this distinction clearer, let us use a standard example: Consider the well-known Müller-Lyer illusion (cf. Müller-Lyer 1889), as shown below (Illustration 2.1).[7]

It contains two arrows whose shafts are of the same length. However, the fins of the two arrows point in different directions. The fins of the one arrow point outwards, while the fins of the other arrow point inwards. This creates the illusion that the shaft of the first arrow is longer than the shaft of the second though both are of the same length. We can convince ourselves that this is true. We can measure the two shafts with a ruler. Interestingly, this will not persuade our eyes. The one shaft still looks longer than the other. In that case, however, we should choose not to trust appearances. Analogously, an act might *seem* wrong, and the claim that it is right may strike us as inadequate. However, on reflection, we may come to believe,

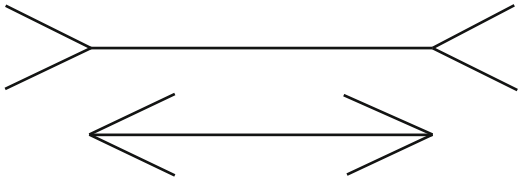

Illustration 2.1 The Müller-Lyer illusion

[5]For a detailed defence, however, see Daniels (1996).

[6]Similarly, Appiah (2008, 113) uses an analogy between reasons for perceptual belief and reasons for action (as opposed to reasons for normative beliefs that we are interested in).

[7]Maria Mukerji suggested this example.

nevertheless, that it is not. That is, we may eventually come to believe that our intuition is unreliable in this instance just like our sense of vision is sometimes inaccurate. It is important to stress, then, that there is a difference between having a moral intuition and adopting this intuition on reflection as a *belief* (cf., e.g., Kagan 2001, 55 and Tännsjö 2011, 307).[8]

This said, we can introduce an approximate criterion of justification for moral doctrines. Rawls says that we may provisionally think of them as the "attempt to describe our moral capacity" (Rawls 1971/1999, 41) and the high-level and low-level moral intuitions that issue from it. This suggests that their acceptability is determined, at least in part, by how well it fits the moral claims which we intuitively endorse. Elsewhere, I called this criterion "intuitive fit" (Mukerji 2013c, 299).

It is evident, though, that this criterion of intuitive fit cannot be the only measure for the acceptability of moral theories. If it were, we would not need them. We should then directly endorse a complex of very specific and unconnected moral principles that just state our moral intuitions. We should, in other words, adopt an "unconnected heap of duties" (McNaughton 1996). Such a construct, however, would not seem very attractive. First of all, it might not even count as a moral theory on the monist interpretation which we have adopted for the purpose of this inquiry. Recall that, according to this interpretation, the theoretical component of a moral theory can be represented as a criterion of rightness. It is doubtful whether a single criterion can be made to fit all our considered moral judgements. Furthermore, a theory like that would not do what we may reasonably expect it to do, viz. "to achieve an acceptable coherence" (Daniels 2011) between our various intuitive judgements that explains and justifies them.

To be sure, by "coherence" we do not mean "coherence with our moral intuitions" (Wood 2008, 47). This requirement is entirely distinct from intuitive fit. It concerns the internal structure of a moral theory, i.e. the relations in which its individual moral claims stand to one another. It does not concern, that is, its external relation to our intuitions. In fact, a fully coherent theory may be one which consists only of highly counter-intuitive claims (cf. Sayre-McCord 1985).

Now, what does the notion of coherence involve?[9] According to a standard interpretation, it requires, first of all, that the claims we endorse not contradict each other. As Rawls points out, there is no reason to suppose that our intuitive moral judgements fulfil this requirement (cf. Rawls 1971/1999, 42). Verifying this is easy. Take two views which seem intuitively appealing. E.g., take the idea that an act which produces the best possible consequences is always right.[10] Moreover, take the view that harming an innocent person is always wrong. Both of these views, I believe, appear, intuitively, quite credible. At any rate, they should appear credible

[8]Note that this sense of "intuition" is different from the sense in which some moral intuitionists have employed the term. They have apparently taken self-evident truth as a defining characteristic of an intuition (cf. Lillehammer 2011, 184).

[9]The notion I work out here has been described as a narrow notion of coherence (cf. Rawls 1974–1975, Daniels 1979).

[10]Even critics of consequentialism have conceded that this idea seems *prima vista* trivially true (cf., e.g., Nida-Rümelin 1993, 1).

to the layman. Then, take a case, e.g., Judith Jarvis Thomson's *Fat Man* case (cf. Thomson 1976, 207–208). Imagine that I am standing on a footbridge over a railway, watching a runaway trolley hurtling down the tracks. I can tell that, if nobody stops the trolley, it will crash into and kill the five people who are working on the tracks. The only way for me to halt the trolley is to push a fat man, who is standing next to me, off the bridge and onto the tracks. Sure enough, this will kill him. However, it will stop the trolley and save the five. Arguably, then, pushing the man has better consequences than not pushing him. It will save a net four lives. According to the first intuition, then, it is right for me to shove the man off the bridge. Yet, since it will also inflict severe harm on an innocent person, it is wrong according to the second intuition. The latter says that it is always wrong to do this. Hence, our pre-theoretical intuitions contradict each other in this case. A moral theory ought to avoid such contradiction. This is the requirement of *consistency*. Insofar as it is a part of the requirement of coherence, it is also a part of the Rawlsian Approach (cf. Kappel 2006, 132).

However, consistency is only a necessary and not a sufficient condition for coherence. The beliefs that $7 + 5 = 12$ and that snow is white are consistent. But there is nothing which *connects* them. Hence, a belief system which contains only those two convictions is not coherent. For coherence requires, secondly, what could be called "systematicity" or "connectedness." (cf. Sayre-McCord 1985) The elements of a moral theory are supposed to systematically link up with one another. This is important to create a sense that the doctrine as a whole is not just an arbitrary collection of randomly assorted components.

Let us consider an example from applied ethics which brings out rather nicely how the criterion of systematicity can be exploited to support a moral claim.[11] Suppose I have the following intuitions about two moral cases.

Intuition 1
If I come across a shallow pond where a child is drowning, and I can save the child at the trivial cost of ruining my best pair of shoes, I ought, morally, to save the child.

Intuition 2
Giving to charity, though it is undoubtedly a good thing to do, is not morally required of me. It is optional.

Singer (2009) points out the following. It is reasonable to assume that, if I give a relatively little amount of money, comparable to the costs of a good pair of shoes, to charity, this suffices to save a child from death by starvation or preventable diseases (e.g. measles, malaria, diarrhoea). Why, then, should it be wrong for me not to save the child in front of me, but permissible not to give to charity? My two intuitions seem to be hard to square. However, what exactly is the problem here? These intuitions are clearly consistent. Apparently, the reason I should be troubled by them lies, then, in their seeming lack of unity. It lies in the fact that my intuitions are entirely disconnected. For this reason, they appear to be arbitrary, and

[11] The example is taken from Singer (1972). We have already used part of it on page 8.

I cannot be sure that they "do not simply express some form of irrational prejudice." (Lillehammer 2011, 176) Evidently, if I would reject Intuition 2 and accept a duty to give to charity instead, I could square this view with my Intuition 1, which says that it is wrong not to save the child. Then, I could bring both my views under, e.g., Singer's proposed *principle of harmprevention*. It says that I ought to prevent a great harm if this costs me comparatively little. I feel a strong inclination, then, to revise my views in the way Singer suggests because I want to make them coherent. This, of course, is precisely the point of the argument. As we can see, then, the systematicity or connectedness of our views can be used as an evaluative criterion besides intuitive fit and consistency.

Let us take stock, then. We have established that the Rawlsian Approach to moral evaluation contains, at least, three sub-criteria. I proposed to call these sub-criteria intuitive fit and coherence. We can factorize the latter into consistency and systematicity or connectedness. In short, then, on the Rawlsian Approach, a moral theory is acceptable to the extent that it is consistent, fits our moral intuitions, and establishes explanatory connections between them.

This idea obviously requires interpretation. Before we proceed by considering various understandings of it, however, let me add a short note on *simplicity* or *economy*. Philosophers often suggest that the acceptability of a theory depends partly on parsimony in its use of fundamental concepts. I propose to disregard this criterion, however – for two reasons. Firstly, the simplicity of a moral doctrine cannot, it seems, make up for its lack of intuitive fit and consistency (cf., e.g., Williams 1973, 137 and 1985, 17; Ross 1930/2002, 23).[12] Secondly, there appears to be a significant correlation between systematicity and simplicity. Theories which contain a dense web of systematic connections between its individual parts tend to possess fewer fundamental concepts than others.[13]

2.2 Interpretations of the Rawlsian Approach

There are various possible interpretations of the Rawlsian Approach. It is a multi-dimensional evaluative criterion. For one thing, then, it is possible to attribute different weights to its distinct sub-criteria. Following Kant's dictum that consistency is a philosopher's greatest duty,[14] this sub-criterion may be seen as a disqualifier. Trade-offs, however, can be made between intuitive fit and connectedness (cf., e.g.,

[12]E.g., Bernard Williams has this to say about simplicity: "If there is such a thing as the truth about the subject matter of ethics (. . .) why is there any expectation that it should be simple? In particular, why should it be conceptually simple, using only one or two ethical concepts, such as *duty* or *good state of affairs*, rather than many? Perhaps we need as many concepts to describe it as we find we need, and no fewer." (Williams 1985, 17; emphasis in the original).

[13]This fact is illustrated, e.g., by Classic Utilitarianism which is often described as both a highly systematic and a rather simple doctrine.

[14]Parfit (2011, xlii) ascribes this *dictum* to Kant.

Kappel 2006, 132). Moral philosophers, of course, differ on the appropriate trade-off ratio.[15] Some theoreticians have vigorously taken the stance that intuitive fit is more important than connectedness. G. E. Moore, e.g., may be interpreted in that way. He says that it is not "the proper business of philosophy, however universally it may have been the practice of philosophers," "[t]o search for 'unity' and 'system' at the expense of truth" (Moore 1903/1959, 222). Tom Nagel has maintained that, "[i]f arguments or systematic theoretical considerations lead to results that seem intuitively not to make sense (. . .), then something is wrong with the argument and more work needs to be done" (Nagel 1991, x). John Stuart Mill and Immanuel Kant famously took the contrary stance. They emphasized the importance of unity and system in moral theory (cf., e.g., Mill 1863, Kant 1785).[16] A further interpretive issue arises in regards to the sub-criterion of intuitive fit. It requires that a moral theory fit the judgements we intuitively endorse. We need to specify this and shall do so in what follows.

There is a consensus, I think, that intuitive fit does not require a moral theory to fit all our intuitions. It merely requires that it match the ones which possess an "initial credibility" (Scheffler 1954, 181). (Ultimately, we may not even demand that it meet all of those since the set of initially credible intuitions may be inconsistent). Rawls says, e.g., that we can discard "those judgments made with hesitation, or in which we have little confidence. Similarly, those given when we are upset or frightened, or when we stand to gain one way or the other can be left aside." It is easy to see why we should dismiss such intuitions as irrelevant. They are dubious from the start. That is, they are not initially credible. We should restrict ourselves, then, to intuitive judgements "rendered under conditions favorable to the exercise of the sense of justice, and therefore in circumstances where the more common excuses and explanations for making a mistake do not obtain" (Rawls 1971/1999, 42). Rawls calls them "considered judgements."

The category of considered judgements is useful to draw attention to the fact that we should not see all moral intuitions as relevant to ethics and that it is important to preselect them before we use them to test moral theories. However, it is by no means clear what makes an intuitive moral judgement a *considered* moral judgement. It is unclear, that is, what distinguishes disliverances of intuition with initial credibility from ones that lack it. Rawls only gives us a few examples. Sure enough, the factors he mentions are plausible. Some emotions are undoubtedly associated with the way we judge moral matters (cf., e.g., Greene et al. 2001; Greene 2008; Huebner, Dwyer, and Hauser 2009), as is self-interest (cf., e.g., Bazerman and Tenbrunsel 2011, 50; Thompson and Loewenstein 1992 and Wright 1996, 13). However, drawing only

[15] To be sure, our talk of a "trade-off" should not be taken too literally. It does not suggest that there are precise, quantitative measures for the overall intuitive fit and systematicity of a moral theory or a single metric on which both can be compared. A moral theory may be evaluated in terms of both its intuitive fit and its systematicity based on a "'seat of the pants' feel." (Putnam 1981, 132) The appropriate trade-off relation, i.e. the overall fit of the theory with our evaluative criteria as a whole, may be determined in the same way.

[16] In fact, both rejected intuitive fit as an evaluative criterion.

on Rawls's ideas, we cannot make progress towards a definite interpretation of the sub-criterion of intuitive fit. So let us look at some views moral philosophers have expressed.

We may distinguish, roughly, between three interpretations of intuitive fit. Each is based on a different second-order theory about the credibility of our moral intuitions. We shall refer to the first as the Top-Down Approach (TD), to the second as the Reflective-Equilibrium Approach (RE) and to the third as the Bottom-Up Approach (BU).[17] To distinguish these three approaches, we need to draw on a differentiation that we made above. Above we discerned low-level intuitions and high-level intuitions. Low-level intuitions, we stipulated, are those intuitions which are less abstract. They concern only one particular case or a very narrow range of cases. In contrast, high-level intuitions are about more abstract moral principles. They cover a broader range of cases or even all possible ones.[18] Now, the distinction is certainly both vague and non-exhaustive.[19] However, I believe that it suffices for the purpose at hand. There are cases in which it seems pretty clear that we are having a low-level or high-level intuition. E.g., if we have an intuitive conviction about whether or not it was wrong for Bill Clinton to lie (or tell a misleading truth) to the American public when asked whether he had sexual relations with Monica Lewinsky, we clearly have a low-level intuition. We make an intuitive judgement which covers only one very specific case. In contrast, if we have the intuition that one ought to act only according to that maxim whereby one can, at the same time, will that it should become a universal law, we clearly have a high-level intuition. Similarly, we certainly have a high-level intuition if we intuitively judge that all sentient beings always deserve to be given equal consideration. Such intuitions cover all possible cases. There are surely moral propositions which lie in between these examples and do not fall clearly on either side of the distinction. This, however, should not be a problem in the present context. With the differentiation between low-level and high-level intuitions in mind, we can define the three interpretations of intuitive fit as follows:

Top-Down Approach (TD)
Only high-level intuitions are initially credible. Hence, moral theories should be judged only by the degree to which they fit our high-level moral intuitions.

[17] There are, of course, innumerable possibilities when it comes to the concrete shape of the respective moral-epistemological theory. For our purposes, however, a rough classification suffices.

[18] Sandberg and Juth (2011) employ a similar distinction between what they call "practical" and "theoretical" intuitions though they draw it in terms of a different criterion. They take them to have different objects. Practical intuitions, they say, are intuitions about cases which is what we call low-level intuitions. Theoretical intuitions are intuitions about moral principles and, apparently, certain metaethical questions too (e.g. the question "what morality is about"). Theoretical intuitions in Sandberg's and Juth's sense should normally be high-level intuitions. There may, however, be instances of intuitions about very specific moral principles which apply only to very few cases. These are, then, theoretical low-level intuitions.

[19] Some authors have distinguished a further category, to wit, *mid*-level principles (cf., e.g., Bell 2007, 71).

Reflective-Equilibrium Approach (RE)
Both high-level and low-level intuitions can be initially credible. Hence, moral theories should be judged in accordance to the overall fit with intuitions both at the high and the low level.

Bottom-Up Approach (BU)
Only our low-level intuitions are initially credible. Moral theories should, hence, be judged only based on the degree to which they fit our low-level intuitions.

To make sense of these approaches, it may be instructive to connect them to the work of some acclaimed philosophers. TD, it seems, can clearly be attributed to the utilitarian philosopher Henry Sidgwick (cf. Singer 1974).[20] In the sixth preface to his legendary book *The Methods of Ethics* (1874/1907), written shortly before his death, Sidgwick provides evidence of this. He explains that he felt forced, at some point, "to recognize the need of a fundamental ethical intuition" (a high-level intuition, as we call it) without which his utilitarian moral philosophy could not "be made coherent and harmonious." (Sidgwick 1907, xvi–xvii)[21]

RE is, as it were, the standard interpretation of intuitive fit. We can attribute it to John Rawls.[22] He describes the process of drawing up a moral theory as a "going back and forth" (Rawls 1971/1999, 18) between the level of the moral principles he seeks to derive, the even more abstract ideas which serve as the premises of this derivation and intuitive convictions about particular cases to which principles are subsequently applied. In doing this, he acknowledges that considerations at all levels of generality – high and low – play a role in the assessment of a moral doctrine.

BU, too, appears to be quite a widespread view. A moral theorist "often starts with intuitions about particular cases and attempts to uncover the general moral principles that underlie these intuitions" (Kahane 2013, 421). Those who favour this approach to theory construction should also hold the corresponding view about theory evaluation. They should hold that a moral theory is acceptable insofar as it implies our low-level intuitions. This view is clearly present, e.g., in works of Phillippa Foot and Judith Jarvis Thomson. They are well-known for their pioneering work on *trolley cases*.[23] These are thought experiments which are designed to trace

[20] Note, however, that this is not the most common reading of Sidgwick. Many philosophers follow Rawls (1971/1999) who, drawing on Schneewind (1963), claims that Sidgwick endorsed RE which is the approach favoured by Rawls himself. I believe that Sidgwick was misinterpreted by Rawls and Schneewind and that the remarks he makes about common-sense morality were falsely taken to represent his own views.

[21] Some may think that Immanuel Kant would also fit the description of TD since he undoubtedly pursued moral philosophy in a top-down fashion. His ambition was to develop a system in which every moral proposition is justified in terms of one supreme principle of morality: the Categorical Imperative. For this reason, some have seen him as a proponent of the TD approach to intuitive fit (cf. Singer 2005). But this would be a mistake since Kant never accepted intuitive fit as an evaluative criterion for moral theories (cf. Nida-Rümelin 2002, 22).

[22] John Rawls explicitly rejects TD to which he refers as the "Cartesian" view. He says that "[t]here is no set of conditions or first principles that can be plausibly claimed to be necessary or definitive of morality and thereby especially suited to carry the burden of justification" (Rawls 1971/1999, 506).

[23] A further major exponent of BU is Frances Kamm. She explicitly states BU in Kamm (2007, 5) and Kamm (1996, 10–12). Interestingly, even the utilitarian philosopher and economist John

out our low-level intuitions to construct high-level moral principles which explain them. We will consider them in more depth below.

We have distinguished the various interpretations of intuitive fit. Now, which one is adequate? Before we address this question, we should note, however, that our argument does not depend on any particular view being *correct*. Rather, the only thing that counts is that one view – viz. TD – is inadequate or, conversely, that either BU or RE is adequate.[24] In what follows, we shall try to establish this by looking at the rationales for each approach.

2.2.1 The Top-Down Approach

Let us look, first of all, at the reasons for TD. Its proponents argue that low-level intuitions are unreliable. They think that we should, hence, rely only on high-level intuitions which they take to be more credible. Their case for TD is based, then, largely on an argument against low-level intuitions. In recent times, proponents of TD have increasingly done this using empirical findings from psychology and related areas. Of course, since this is not a tract in moral epistemology, we can only look at a few examples.[25]

Harsanyi has, at times, made remarks that may be read as expressing a sympathetic attitude towards BU: "Should the axioms of my ethical theory turn out to possess morally unacceptable practical implications," he says, "(. . .) then I must be always willing to revise my axioms" (Harsanyi 1977b, 26).

[24]For an argument to that effect, see Mukerji (2014).

[25]It should be noted that many philosophers regard empirical considerations as beside the point when it comes to the evaluation of moral theories. Drawing on well-known ideas predominantly by Hume (1888/1960) and Moore (1903/1959), they argue that there is a metaphysical divide between the spheres of 'Is' and 'Ought' and that moral facts are distinct from and not definable in terms of natural facts. Ethics, they say, is hence *autonomous* in the sense that no empirical facts could conceivably influence the moral question whether a given action is right or whether a given moral theory is adequate. In reply to such concerns, it should be stressed that empirical arguments do not generally claim that normative propositions follow straightforwardly from empirical propositions (Mukerji 2015). Hume's and Moore's points are usually conceded. It is claimed, however, that certain information about the workings of our moral faculty may be useful when it comes to figuring out which principles are justified. But their justification may itself be independent from empirical matters. An analogy may be useful to drive home the point. Consider our visual sense. It is normally reliable and helps us to figure out what goes on in the world around us. But there are optical illusions (e.g. the Müller-Lyer illusion that we considered on page 20). We should, therefore, be interested in understanding the conditions under which these illusions arise. For, when they obtain, we are, it seems, well advised to put less faith in our visual perceptions than we normally do. Similarly, we may believe that our moral sense is normally reliable. But there may be certain facts about it that cast doubt on its judgements in certain situations.

One obvious requirement for the reliability of an intuition is that it passes Sidgwick's "criterion of consent."[26] (Sidgwick 1879, 108) He argued that

> the denial by another of a proposition that I have affirmed has a tendency to impair my confidence in its validity. (. . .) For if I find any of my judgments, intuitive or inferential, in direct conflict with a judgment of some other mind, there must be error somewhere: and if I have no more reason to suspect error in the other mind than in my own, reflective comparison between the two judgments necessarily reduces me temporarily to a state of neutrality. (Sidgwick 1907, 341–342)[27]

In other words, if we encounter a reasonable person who disagrees with us about some moral question, this should decrease the confidence that we are right. (We need not even go so far as to become completely neutral, as Sidgwick suggests.) Now, the question is whether people seem to disagree comparatively more about low-level matters. There is one difference between low-level and high-level intuitions which may suggest that there must be more disagreement regarding low-level intuitions. One could claim that differences about low-level intuitions are much more likely than opposing intuitive views at the higher level since there are innumerable cases at the low level while there is a limited class of principles which cover all cases. So it is much more likely that we will ever reach agreement on a confined set of high-level judgements than about a potentially infinite amount of convictions about particular cases. A further point one could make is that people disagree regarding their low-level intuitions to quite a large extent.

The second reason one might give preference to high-level intuitions has to do with what psychologists call "framing effects." Let us, first of all, consider what these effects are. In an oft-cited paper, Amos Tversky and Daniel Kahneman report that people's intuitions about how one should act in particular cases may change depending on how the choice is verbally framed.[28] They had two groups of participants face a decision problem between two policies A and B and C and D, respectively. The description of the case was as follows:

> Imagine that the U.S. is preparing for an outbreak of an unusual Asian disease which is expected to kill 600 people. Two alternative programs to fight the disease, A and B, have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows: (Tversky and Kahneman 1981, 453)

The first group got this description of A's and B's consequences.

> If program A is adopted, 200 people will be saved. If program B is adopted, there is a 1/3 probability that 600 people will be saved, and a 2/3 probability that no people will be saved. (Tversky and Kahneman 1981, 453)

[26] Note, however, that the criterion of consent is not universally accepted (cf., e.g., Smart 1956, 346).

[27] This point has been made by other authors, e.g. by Ross (1939, 88).

[28] It should be noted that philosophers have recognized the existence of framing effects before psychologists did. Williams (1970), e.g., describes the phenomenon in the context of the issue of personal identity and suggests that intuitions about a given scenario can differ under two equivalent descriptions.

The second group was given a choice between C and D (instead of A and B). This was the description of their consequences:

> If program C is adopted, 400 people will die. If program D is adopted, there is a 1/3 probability that nobody will die and a 2/3 probability that 600 people will die. (Tversky and Kahneman 1981, 453)

Each group stated their preferences about the programmes.[29] In the first group, 72 % favoured option A and 28 % option B. In the second group, 22 % preferred option C and 78 % option D. Note that the only information participants had about programmes A, B, C, and D was regarding their effects. Programmes A and C and programmes B and D, respectively, had the same effects though these were framed differently (i.e. A and B regarding lives *saved*, C and D relating to lives *lost*). Such a difference in verbal framing, it seems, should not make a difference in the moral evaluation of the respective options. However, as Tversky and Kahneman (1981) showed, it does appear to affect people's judgement.[30]

Now, why should framing effects speak against low-level intuitions and for high-level intuitions? This is because lower-level intuitions seem particularly susceptible to framing effects while higher-level intuitions appear to be comparatively immune to them. This suspicion, one may argue, is supported by the fact that hardly any research shows that framing effects exist in high-level intuitions, whereas there is plenty of evidence which suggests that low-level intuitions change due to framing (e.g. Tversky and Kahneman 1981; Petrinovich and O'Neill 1996; Haidt and Baron 1996).

A third reason to be sceptical of low-level intuitions *vis-a-vis* high-level intuitions is that there might be *debunking explanations* for intuitive judgements at the low level, but not at the high level. In particular, the fact that we have certain low-level intuitions might be because we have genetic or cultural dispositions for emotional responses to specific cases. Take, e.g., our intuitions about killing. The neuropsychologist and philosopher Joshua Greene and his colleagues conducted a series of experiments to study how firmly and under which conditions we disapprove of killing an innocent person. Amongst other things, they compared a pair of cases which they called *Footbridge* and *Remote Footbridge*, respectively. In *Footbridge*, participants were supposed to imagine a situation in which a runaway trolley threatens to kill five people who are working on the tracks of a railway. The only chance

[29] To be sure, participants were not asked specifically for their moral intuitions in these cases. But since their own self-interest was not at stake in either case, it is reasonable to suppose that their favoured choice is based on moral considerations only (cf. Sinnott-Armstrong 2008, 55).

[30] There are various types of framing effects and not all of them depend on the wording of cases. A quite famous sequence-related framing effect is associated with the work of the philosopher Peter Unger. To discredit intuitions about cases Unger (1996) remarks that in moral thought experiments – most notably the "trolley cases" due to Foot (1978, 19–32) and Thomson (1976, 1985) – there are usually only two choice options. And we tend to have strong intuitions for or against, respectively, one of the options. Using his "Method of Several Options," he attempts to show that our intuitions about which option is right change as we add further options to the choice problem.

to save them from the approaching trolley is to push a fat man off the footbridge over the tracks, thus using him as a trolley stopper. This would kill him. However, it would save five lives. The *Remote Footbridge* case involves basically the same scenario, except that this time the only chance to save the five is to hit a switch. This switch will open a trapdoor on which the fat man is standing. He will drop onto the tracks and will stop the trolley. Once more, this will kill him. However, it will also save the five (as does pushing the fat man in the previous case). Most participants judged that pushing the man in *Footbridge* is wrong. Significantly fewer people, however, had the intuition that it is wrong to kill someone by hitting the switch in *Remote Footbridge* (cf. Greene et al. 2009).[31] Greene's explanation for this is that our species has, as a matter of contingent fact, developed emotional "point-and-shoot" responses to types of cases that our ancestors frequently faced. We have an aversion to anything that feels like applying force in an up-close and personal manner. Now, "[t]he thought of pushing the stranger off the footbridge elicits these emotionally based responses" (Singer 2005, 348), while the notion of hitting a switch, which essentially produces the same effect, does not.[32] Given that "these moral intuitions are the biological residue of our evolutionary history," however, "it is not clear why we should regard them as having any normative force."[33] (Singer 2005, 331)

The important point to be added to this is that we apparently cannot say the same about our high-level intuitions. They are abstract and do not elicit the contingent emotional responses which we can explain (away?) by our evolutionary history. They are not, as Singer claims, intuitions in the ordinary sense, but rather "rational intuitions" (Singer 2005, 351) and, hence, more trustworthy and relevant for the assessment of our moral doctrines.

2.2.2 The Reflective-Equilibrium Approach

Now, what can those who support the other approaches say in reply to the above? Both proponents of RE and BU need to show that at least some low-level intuitions seem initially credible and that the justification of a moral theory should, hence, be assessed partly or entirely regarding its fit with low-level intuitions. How can this be done?

[31] Participants were asked to report on a 9-point scale how strongly they approved/disapproved of the killing. Killing the man in the footbridge case received an average rating of 3.89 (standard error 0.22). Killing the man in the remote footbridge case received an average approval of 5.14 (standard error 0.20).

[32] Indeed, the claim that many of our intuitive judgements about particular cases are based on emotional responses is confirmed by a number of recent neuroimaging studies performed by the psychologist and philosopher Joshua Greene (cf. Greene 2008). Greene showed that certain judgements about cases are associated with increased neural activity in emotion-related areas of the brain (e.g. posterior cingulate cortex, medial prefrontal cortex, amygdala).

[33] Crisp (2006, 24) makes essentially the same point.

Let us reconsider the issue of interpersonal variation. Above, we said that there were innumerable cases. Hence, there seems to be a much greater potential for disagreement about low-level intuitions. One may corroborate this impression using findings in psychology and the social sciences which report wide-ranging controversies about cases. Now, it has to be conceded, of course, that whenever we find that other people have different intuitions we are well advised to follow Sidgwick's advice and take a sceptical attitude towards our own intuitive leanings. However, this does not mean that we should be sceptical about all our low-level intuitions. Some philosophers have argued – I believe rightly – that we should take empirical findings regarding intuitive disagreements on cases with a pinch of salt. The reason is that there seems to be a selection process at work. Neuro-psychologists like Joshua Greene purposefully select cases people tend to disagree on because they want to explain the differences in their judgements, e.g. regarding the differences in their brain activities (cf. Greene et al. 2001). The wide range of cases in which our low-level intuitions coincide is not as interesting and is not reported as frequently. Hence, when we read such studies, we get the impression that there is disagreement about almost each and every case. Bernard Gert makes the same point. He says that moral questions "such as whether it is morally acceptable to hurt someone simply because you dislike him are not controversial at all, but because they generate no discussion they tend to be forgotten." (Gert 2004, 14).

Let me add a second point which is surely worth stressing. When we survey the philosophical literature, we come across many cases that apparently exhibit strong disagreement. From this, too, we might conclude that there is probably much disagreement about cases in general and that, therefore, our low-level intuitions are to be doubted as well. However, this would, again, be an inference from a biased sample. Many cases in moral philosophy (and elsewhere in philosophy) have the purpose of drawing out the implications of competing theories and testing them against our intuitions (cf. Dennett 1984, 17–18). They often serve as the basis for *reductio* arguments. One theorist says to another: "Let us assume that your theory is correct. This would mean that in case X it would be right to do Y. But, surely, that view is absurd." To defend her theory against such an objection, the other theorist can either deny that it implies act Y in case X. Or she can simply embrace it and say: "I find doing Y in case X very reasonable." It is easy to explain, then, why there would be such a great deal of disagreement about cases in moral philosophy. The (reported) intuitions of theorists about cases differ, at least partly, because these intuitions have the function to attack competing doctrines and to corroborate one's own theoretical stance.

If this is in fact so, much of the disagreement about cases – at least amongst professional philosophers – is explained by their theoretical disputes over the right moral theory.[34] This, in turn, suggests that there should be a roughly proportional

[34]I am indebted to Martin Rechenauer who suggested to me that disagreement about cases may be seen as the embodiment of a more deeply rooted disagreement about moral principles. This point also seems to be acknowledged by Norcross (2008, 66) who says that he is "all too aware that

amount of disagreement about high-level intuitions since moral doctrines are high-level matters. If cases in moral philosophy do, in fact, mainly serve the purpose of putting a high-level principle to the test, there should be roughly one such principle for each case on which philosophers disagree. We can, of course, multiply these cases by specifying morally insignificant details differently. However, if they serve the same theoretical purpose, I would suggest to view them as essentially the same case.[35]

Now, let us turn to high-level intuitions on which there is allegedly comparatively little disagreement. This is simply not true. There is much disagreement. Take, e.g., Peter Singer's principle of harm prevention which we considered previously. He claims that "if it is in our power to prevent something very bad from happening, without thereby sacrificing anything morally significant, we ought, morally, to do it." (Singer 1972, 231) This principle implies that I ought to save a child who is drowning right before my eyes. However, it also entails that I ought to prevent a child from dying in some remote place in Africa if I can do this at roughly the same costs (e.g. by donating money). The principle contains no reference to physical distance – and rightly so, finds, e.g., Unger (1996). Frances Kamm famously disagreed with this and claimed that "at least intuitively, distance per se matters to what obligations we have." (Kamm 2007, 352) A further example is the high-level intuition that, we ought, *ceteris paribus*, to prevent a greater rather than a smaller harm if we cannot prevent both. E.g., if a flood is threatening the lives of people on both sides of an island and I am the captain of a freight ship who can save people on either side, but not on both sides, I ought, morally, to act so as to rescue more people and prevent the greater harm.[36] This, I take it, sounds plausible to most people. There is, however, no universal agreement about this case. Some philosophers have disputed it (e.g. Taurek 1977; Lübbe 2008). It appears, then, that the seemingly larger disagreement about low-level intuitions does not discredit them to a greater extent than high-level intuitions.[37]

---

non-consequentialists' intuitions diverge radically from [his] own" consequentialist intuitions and by Prinz (2010, 387) who remarks that "[p]hilosophers intuitions are not theory-neutral" which, as he hypothesizes, may be "one reason why philosophers seem to have different intuitions about the same cases."

[35]My view is corroborated by a remark by Shelly Kagan. He makes the point that "typically when we think about cases, we are only thinking about *kinds* of cases." (Kagan 2001, 61–62; emphasis in the original) This suggests that different specifications of a case structure can be seen as the same case. On the same point, see also Appiah (2008, 84–85).

[36]This example is taken from Taurek (1977).

[37]One might, however, draw a generally sceptical lesson from all of this and conclude that neither low-level nor high-level intuitions are reliable (cf. Singer 2005, 349). I find this hardly plausible. In many areas, our intuitions are very unreliable in isolation. But we would not conclude from this that we cannot make progress in these areas. Consider, e.g., probability theory. Many simple card tricks are able to fool us because our intuitions about probabilities are very unreliable. Nevertheless, human beings were able to develop a probability calculus based on intuitive considerations which, over time, got more and more formalized. And this probability calculus is very reliable, e.g., in

Above we said that empirical results concerning our moral intuitions and, notably, findings regarding their evolutionary genesis may cast doubt on them and, in particular, on low-level intuitions. What can we say in reply to this? There seem to be three strategies to defend low-level intuitions against these charges.

- Firstly, there is the strategy of blunt denial. We may say that the empirical findings which adherents of TD use to discredit low-level intuitions are just irrelevant in the context of moral theory.
- Secondly, the empirical results themselves can be challenged.
- Thirdly, empirical findings can be acknowledged, but the link between these findings and the conclusion drawn by proponents of TD can be disputed.[38]

Those who opt for the first strategy may defend their view by pointing towards Hume's (1888/1960) crucial distinction between *Is* and *Ought*.[39] They can claim that, as a matter of principle, it is not possible to draw conclusions for moral theory from factual evidence and that, therefore, the above considerations are fallacious.[40] This would, of course, be an argument against a straw man. No serious thinker would suppose that we can infer normative conclusions straightforwardly from empirical evidence (Mukerji 2015). Rather, those who claim that facts about our low-level intuitions ought to make us suspicious as to their credibility. This is a moral-epistemological claim which does not derive from any fact. It is quite a plausible claim, too! Facts about our moral psychology obviously matter. In this connection, John Rawls may serve as a crown witness. As we saw above, he thinks we should be wary of intuitive judgements made when we are upset, frightened, or stand to gain one way or the other because they are likely to be distorted. In saying this, he plainly acknowledges the relevance of empirical psychology to moral theory.[41]

---

predicting the frequency of future events. Since we are not concerned with the issue of scepticism we can put this issue aside. An instructive overview over sceptical positions on ethics can be found in Sinnott-Armstrong (2006).

[38]These three possible replies are inspired by Timmons (2008, 93) suggestions as to how a deontologist can defend herself against Greene's attack on their theory.

[39]Another possible strategy is to argue that the capacity for ethical intuition is an *a priori* faculty (cf. Lillehammer 2011, 176). It is very unlikely to work. So, for reasons of scope, we shall put it aside.

[40]It should be noted, however, that the precise purport of the Humean thesis is unclear. At least on some interpretations, it is clearly false, as Prior (1960) has shown. Consider a factual proposition E. From E we can derive $E \vee N$, where N is a normative proposition. There are only two possibilities. $E \vee N$ is a normative proposition. In that case, an *Ought*-proposition can be derived from an *Is*-proposition. Or it is factual. In that case, however, it is possible to derive N from $\neg E$ and $E \vee N$ which are factual propositions. So Ought-propositions can be derived from Is-propositions in any case. For a thoroughgoing treatment of the problem identified by Prior, see Schurz (1997).

[41]In recent times, the exponents of a new movement in Philosophy called "Experimental Philosophy" have vigorously defended the relevance of empirical data for philosophical theories (Knobe and Nichols 2008). On the relevance of psychological findings for moral philosophy, see also Driver and Loeb (2008), Greene (2008), and Prinz (2010).

The first strategy is off the table, then. The second strategy would fall under the purview of an empirical scientist. Therefore, it, too, is off the table – at least as far as our present inquiry is concerned. This leaves us with the third strategy, viz. to argue that low-level intuitions are not entirely discredited by empirical findings. How can this be done? Consider, first, framing effects. Does the fact that our low-level intuitions are subject to framing effects show that we should dismiss them *tout court* and trust only high-level intuitions? There are two reasons, I believe, why this would be a hasty conclusion to draw.

First of all, it has not been shown (nor do we have much reason to suspect) that all case-based intuitions are susceptible to these effects. Rather, it has been reported by some researchers that certain experiments could not demonstrate the existence of framing effects. Petrinovic and O'Neill (1996), e.g., failed to detect wording-related framing effects in some cases. To be sure, this does not demonstrate that there were no framing effects. However, it does give us reason to doubt the sweeping conclusion that all our low-level intuitions are susceptible to these effects.

Secondly, even if all low-intuitions were, in fact, affected by framing effects, this would not mean that we have to dismiss them *tout court*. Presumably, framing effects arise from the fact that, e.g., different wordings or different contexts draw our attention to particular features of it. This, in turn, may lead to a well-known problem, to wit, that we neglect other features which may be of equal importance (cf. Brink 1984, 117). If we know that we tend to have such "blind spots," as Bazerman and Tenbrunsel (2011) and Sorensen (1998, 273) call them, we can, it seems, discipline ourselves. We can try to focus on all relevant aspects of a moral problem and carefully consider our intuitive verdicts.

How do we answer the third challenge, viz. that low-level intuitive responses to cases are based on historically contingent emotions? The first line of defence that is possible to launch is to emphasize that the relationship between moral judgements and emotions allows of various interpretations. Even though it might be possible to show that certain emotions accompany certain intuitive judgements, this does not warrant the conclusion that emotions *cause* these judgements. This is "because it by no means follows when two phenomena accompany each other in their variations, that the one is cause and the other effect."[42] (Mill 1882, 496) There are various other possibilities (cf., e.g., Mukerji 2013a, 118–119). To justify the conclusion that variations in some empirical phenomenon x cause changes in another event y, it has to be ruled out, in particular, that

(i) the correlation between these variations is accidental,
(ii) variations in y cause variations in x (rather than *vice versa*),
(iii) variations in some other factor, z, cause variations in both x and y and
(iv) variations in x cause variations in y through some intermediary factor w.

Now, presumably, it can be empirically established that (i) is very unlikely at least when it comes to certain emotions and certain intuitions about cases. Let us

[42]This diagnosis is confirmed, e.g., by Huebner et al. (2009, 4).

assume that empirical scientists did their homework and that they took good care to rule out coincidence using standard methods of statistical testing.[43] However, based on a literature survey, not all of the other possibilities can be ruled out at this stage. There are models of moral reasoning which do not conclude that emotion drives intuitive moral judgements. Instead, they hold that (ii) is true (e.g. Dwyer 1999; Hauser 2008; Mikhail 2007). Scientists who subscribe to such models believe that the direction of causality goes from moral intuition to emotion. According to them, the fact that we feel a certain way about a particular action (e.g. the fact that we are repulsed by the idea of pushing the fat man off the bridge) can be explained by the fact that we have a certain intuition about the wrongness of this act. As reported by Huebner et al. (2009), models of the Piaget/Kohlberg tradition assume what they regard as a Kantian picture of moral judgement. Kant thought, as is well known, that reason gives rise both to the rational emotion of "reverence" for the moral law and the particular moral judgements about cases. Models which adopt this picture support the alternative explanation (iii). They hold, that is, that there is a third factor whose workings determine both variations in emotions and moral intuitions.[44] As I said above, we cannot assess how the respective models, in fact, stand up to empirical evidence. This is a primarily scientific and not a philosophical issue. Scientists have to work out which account is adequate. However, until there is no significant agreement on the issue, we should not make the mistake and listen to just one side of the debate. Hence, we should not jump to the conclusion that our contingently evolved emotional responses to individual cases drive our low-level intuitions, thereby discrediting them.

What is more, it would not even follow that we have to mistrust all our low-level intuitions, even if it did turn out that they are all driven by emotions. To be sure, in many instances the fact that emotion drives an intuition should make us cautious. However, the reason for this seems to lie in the fact that strong emotions have a particular kind of effect. They "cloud" our judgement, one might say.[45] Professional philosophers who have talked with laypeople about moral-philosophical issues can surely confirm this. Sometimes when we ask them to imagine certain abhorrent cases, e.g. cases involving footbridges and fat men, our interlocutors may find the notion of doing a particular act so repulsive that this blinds them to other important factors about the case. Even if this is true, though, it does not follow that we can make the hasty generalization that emotion clouds all our low-level intuitions and that the latter are full of blind spots (cf. Sinnott-Armstrong 2006, 194). To avoid this

[43] Berker (2009), however, has questioned the research reported by Greene (2008) in that way. For a rejoinder, see Greene (2010).

[44] It is, admittedly, quite a stretch to associate this view of moral reasoning with Kant since Kant's moral system is purely based on reason and does not allow moral intuition any role to play (cf. Kant 1785).

[45] For this reason, Sinnott-Armstrong (2006, 194) moots the principle that we need an independent reason to believe intuitions that we have in situations where we are "emotional in a way that clouds judgment." This caveat is important. The principle advises caution only when it comes to emotions which cloud our judgement. And these might not be all emotions.

effect, it seems we just need to make sure that our emotions do not get the better of us and lead us to a judgement that is too brisk. Contrary to inclination, we must ensure that we do not only consider a particular aspect of the case but examine it for all factors which, on reflection, ought to be seen as relevant. If we do this, I see no reason not to trust our low-level intuitions, even if it should turn out that they are laced with emotion. As Tersman (2008) notes, there might even be a reason to think that certain types of emotional involvement might even improve our intuitive verdicts. He argues, e.g., that a "well-founded evaluation of a moral dilemma usually requires information about which interests are at stake, and in order to gather such information it may help if we are capable of some amount of empathy." (Tersman 2008, 393)

As a final note, it may be mentioned that Tersman (2008) also points out that explanations for the genesis of our high-level intuitions are also available. He says that "[a]lready from the start, Christian ethics involved the belief that many differences that had previously been regarded as morally relevant, such as ethnicity or differences in class, are not in fact so." (Tersman 2008, 401) This might explain why Westerners whose societies are coined by a Christian tradition find it so intuitive that all morally relevant subjects deserve the same moral consideration. This is one of the high-level intuitions on which, e.g., Henry Sidgwick bases his utilitarian theory. The same holds for Peter Singer's philosophy. Now, if we can generally regard genetic explanations as casting doubt on our intuitions, they would certainly cast doubt on high-level intuitions, too. It is important to note that we can interpret this reasoning in two ways. One way of interpreting it is as a *Tu Quoque*. In that case, it would not lend support to low-level intuitions. However, coming from a proponent of RE this is not how we should make sense of it. It seems we should rather interpret it as a companionship-in-guilt argument (cf. Mackie 1977, 39). Those who believe in the RE approach believe that high-level intuitions *can* be reliable. When they point out that high-level intuitions may be shaped by tradition, they do not mean to claim, therefore, that this makes them *ipso facto* unreliable. Rather, assuming that high-level intuitions are reliable, they want to point out that proponents of TD are inconsistent when they criticize our low-level intuitions based on their causal history. After all, we could make the same point about the high-level intuitions whose reliability they leave unquestioned. In other words, high-level and low-level intuitions are "companions in guilt." There is no reason to be sceptical about low-level intuitions in particular.

2.2.3 The Bottom-Up Approach

Let us take stock of where we are. In Sect. 2.2.1, we considered the case for TD. We looked at some of the reasons why one might think that a moral theory ought to fit only our high-level intuitions. In Sect. 2.2.2, then, we examined how one could make a case for RE. Proponents of RE, such as John Rawls, think that a moral theory ought to fit our intuitions at both ends. It ought to fit, that is, the relevant intuitions

of high and low degrees of generality and abstractness. Such theorists have to argue that there is no reason to mistrust all our low-level intuitions and to generally give preference to high-level intuitions. As we saw, they can make a persuasive case. For it seems that the arguments presented by proponents of TD, who attack the credibility of low-level intuitions, are rather shaky. Now, supporters of RE reply to the criticisms of the adherents of TD in a rather defensive way. They only seek to establish that there is no reason to discard all low-level intuitions and that moral theories should be tested against them, too. They do not claim, as we just saw, that high-level intuitions are altogether unreliable. For they believe that at least certain high-level intuitions do possess initial credibility. This is where champions of the BU approach come in. They share with those who favour RE the view that we should regard at least some low-level intuitions as initially credible. So they can adopt the case that proponents of RE make in defence of low-level intuitions. They merely need to add to it a criticism of high-level intuitions which shows that the latter do not, in fact, possess initial credibility.

As I said above, nothing in our argument depends on BU being correct. For the purpose of our inquiry, it is sufficient to show that we should reject TD because at least some low-level intuitions possess initial credibility. Both proponents of RE and BU hold this view. Therefore, it is, in fact, unnecessary for us to argue for BU and against RE. Nevertheless, let us, for the sake of sportsmanship, quickly point out why BU might be plausible.

It seems that supporters of BU could say something about high-level moral principles which is similar to what David Hume said about abstract ideas (cf. Hume 1888/1960, 25–33). As an empiricist, Hume believed that all ideas are derived from prior sense impressions. Since every impression is an impression of a concrete object, all ideas, he thought, had to be concrete as well. Thus, Hume reasoned, when we appear to think abstractly, we actually have a concrete idea in mind which we then allow to relate to other objects that are sufficiently similar in its qualities. Something like this may be going on when we think of an abstract principle and form an intuition about it. It may be that we do not consider it in its abstractness, but imagine concrete cases to which it applies and then say "yes" or "no" to it depending on whether its implications in these cases seem intuitively acceptable. It may be, that is, that whenever we think we have a high-level intuition about a principle, we have, in fact, one or more (muddled) low-level intuitions about the anticipated implications of the principle in particular cases that we happen to think up. In support of this thesis, one could, e.g., cite Amartya Sen, who said something remarkable in the context of social choice theory. Social choice theory offers an axiomatic take on moral problems. Most of what happens in it happens on a rather abstract plain, where theorists give much attention to the credibility of the axioms which are more or less formalized versions of high-level moral principles.[46]

[46]Such comparisons are necessary, in particular, when it comes to "impossibility results" (e.g. Arrow 1951/1963), which show that certain axioms cannot logically co-exist. In that case, the theorist has to drop at least one to ensure consistency.

It seems, then, that social choice theory is a paradigm example of TD and that those who practice it should believe that the initial credibility of the axioms plays a great role. Now, curiously, Sen has claimed that "[w]hen we say 'yes' to an axiom we do not think absolutely abstractly. We think of actual cases." (Sen 2009) It seems we can interpret this in the way I just suggested, viz. as saying that the justification of a high-level axiom depends entirely on whether or not its low-level implications intuitively make sense. This, in turn, would suggest that there are, in fact, no high-level intuitions about moral principles. It would mean that they are mere chimeras and should play no role in moral inquiry.

Having said this, allow me, briefly, to draw out what it would mean for the evaluation of moral theories if we adopted the BU approach. It may seem that, on BU, we would always have to talk about cases and would have to eschew any mention of general principles. But this is not so. BU does not suggest that philosophers should entirely disregard the intuitive appeal of principles. It would still allow us to endorse or reject moral theories in light of their compatibility or incompatibility with abstract tenets that we find intuitively plausible. However, it would remind us that when we do this, we must not forget that the intuitive appeal of principles derives from the intuitiveness of its case implications. Since many principles apply potentially to an infinite amount of cases, this suggests that we must always consider the possibility that the intuitive appeal of a principle may, on reflection, turn out to be smaller than it initially appeared. We may discover that a seemingly plausible principle has very counter-intuitive implications in particular circumstances. And this may completely destroy its credentials from the standpoint of intuitive fit on the BU interpretation.

As a final note, it should be stressed that, even if BU is accepted, it does not follow that one should reject high-level principles whenever their implications contradict low-level intuitions. As we worked out above, this is because intuitive fit is merely a *sub*-criterion of the Rawlsian Approach. Within that approach, systematicity may play a great role, too. There can, hence, be a trade-off. Even if principles leave something to be desired regarding their low-level intuitive fit, we may still accept them due to their systematizing strength.

2.3 Provisional Fixed Points

Above we factorized the Rawlsian Approach to theory evaluation into distinct sub-criteria. And we discussed various interpretations of it. Now we need to consider how we can use the approach to develop a workable method for our evaluation of consequentialism. This is necessary since nothing that we have said so far can be straightforwardly applied. This may not be obvious. So let me explain.

The Rawlsian Approach does not, as it were, provide a "pass-or-fail test" for moral doctrines. Hence, we cannot directly apply it to assess consequentialism. The approach offers, rather, a "philosophical ideal" which, presumably, none of our moral doctrines can achieve. We should not, therefore, dismiss a theory, if it does

not attain a perfect fit with the relevant moral intuitions. We should, rather, assess it regarding whether or not it "moves us closer to the philosophical ideal" (Rawls 1971/1999, 43) than its alternatives. However, that would mean that we have to examine not only consequentialism but also its main competitors in what could be called a *comparative study* (cf. Sinnott-Armstrong 2011). This is something which we cannot do here. Given the scope of this inquiry, there is simply no way that we can compare consequentialism even with its most prominent rivals.

There are, however, ways to circumnavigate this problem. We can look for decisive tests that follow from the Rawlsian Approach. As we said above, consistency which is one of its sub-criteria possesses the status of a knock-out criterion. Hence, on the assumption that at least some moral doctrines are actually consistent, we can reject those which are not. For, according to the Rawlsian Approach, they will certainly be inferior to any consistent moral doctrine. That means, if we could show that all consequentialist theories are, indeed, inconsistent, we could conclude that they fail. Certain theorists have, in fact, discussed whether this line of argument can be successful. Some of them have focused on the issue of "complex acts."[47] (e.g. Bergström 1966, Bykvist 2002; Castaneda 1968; Carlson 1999a) Others have brought up charges of self-defeat. They have argued that consequentialist agents fail to achieve aims that are deemed desirable by the lights of their own moral theory. This strategy has been applied, e.g., by Hodgson (1967) and Nida-Rümelin (1993). I shall propose, however, to put it aside here.[48]

Can we try to use coherence? I believe that this would be a bad idea. Such an approach would probably not give us enough to chew on. As some philosophers have pointed out, consequentialist theories are, in fact, rather "unlikely to encounter problems of coherence." (Sumner 1987, 173)

The strategy that seems to fit our present purpose best is based on the sub-criterion of intuitive fit. It uses what John Rawls calls "provisional fixed points" for moral theorizing.[49] I shall, therefore, refer to it as the Provisional Fixed Point Approach (PFPA). The idea behind it is as follows. When we assess a moral theory, we look for intuitive convictions which possess a high degree of initial credibility. They have to be so strong that it seems very reasonable to expect that an acceptable moral theory should fit them. Then, we check whether the doctrine in question does, in fact, match these intuitive judgements. If not, we reject it, no matter how coherent it seems and irrespective of how intuitive it is in other regards.

In the context of our discussion, this approach seems appealing for two reasons. Firstly, we do not need to conduct a comparative study. We do not have to consider the merits and demerits of consequentialism in comparison to its alternatives. PFPA

[47] Elsewhere, I have briefly discussed this strategy (cf. Mukerji 2013c, 306).

[48] It can be argued that at least the second strategy runs into difficulties when certain types of agent-relative consequentialist theories are taken into consideration (cf. Mukerji 2013a, 114–117).

[49] It seems that PFPA is implicitly recognized in many moral-philosophical tracts. In addition, there is a number of authors who have emphasized that the approach plays a great role in the practical application of the Rawlsian Approach. See, e.g., Daniels (1996, 28), Mulgan (2007, 58), Nida-Rümelin (2002, 34–35), Otsuka (2006, 110) and Rawls (1971/1999, 18).

allows us to devote our full attention to the object of our inquiry. Secondly, the approach homes in on what seems to be the most important aspect of the debate about consequentialism. Most critical studies have focused on intuitive fit and have attempted to demonstrate that consequentialism is unacceptably counter-intuitive.

It should be noted, however, that these advantages come at a cost. First of all, since PFPA is based exclusively on one sub-criterion of the Rawlsian Approach, it will miss objections that draw on the other evaluative criteria. Those may turn out to be crucial. Secondly, the assumption assumes that a consistent moral theory can, in fact, fit the respective provisional fixed points. Hence, we have to qualify the conclusions we draw from it with a *proviso*: Should it turn out that it is, in fact, impossible to match the respective provisional fixed points, we have to revoke our verdict.[50] Thirdly, PFPA can only tell us whether we have sufficient reason to *reject* a given moral doctrine. However, it cannot tell us whether we have sufficient reason to *accept* it. It is easy to see why. If we find that a given moral theory violates certain provisional fixed points, we can judge that it ought to be rejected (under the mentioned *proviso*, of course). If we find that a moral theory fits all our provisional fixed points, we cannot judge, however, that it ought, therefore, to be accepted. It might still be possible that the doctrine is, in fact, untenable. All we can say, based on PFPA, is that we do not have reason to think so.

Before we move on, let me briefly address these worries. The first problem naturally arises for any in-depth investigation of a philosophical problem. Granted, in using PFPA, we may lose sight of certain issues which are undoubtedly important in their own right. This seems defensible, however, because we have to confine the scope of the inquiry to ensure its tractability. In reply to the second problem, we can give a similar answer. It is true that, in using PFPA, we do rely on the assumption that it is, in fact, logically possible for a moral doctrine to fit the respective fixed points in question. We cannot ensure that this assumption is justified – at least not within the scope of the present inquiry. However, every philosophical investigation has to take certain things for granted. It cannot address all problems at once.[51] The third problem, I believe, is one we can indeed ignore, given the rather modest aim of the investigation. We are merely interested in developing a case *against* consequentialism. We are not seeking to investigate whether a constructive case *for* consequentialism is possible. With this in mind, it seems that PFPA is adequate for the purpose at hand.

Now that we have, I hope, a clear enough idea about PFPA in the abstract, we should specify it further. In particular, we should answer the question where we

[50] Social choice theory has shown that weak seeming moral judgements may turn out to be logically incompatible. An example which illustrates this is an impossibility theorem proved by Sen (1970b). It is called the "Impossibility of the Paretian Liberal" and shows that a minimal notion of individual rights is incompatible with the Weak Pareto Principle.

[51] In addition, it might be mentioned that the possibility of inconsistency seems to be confined to abstract level fixed points. Fixed points about cases cannot be inconsistent unless they concern the same case. As will become clear in Sect. 6.1, our argument relies entirely on fixed points about cases.

may find provisional fixed points. This gives us a chance to tie up loose ends and to relate what we just said with the points we made in the previous section. Obviously, the answer depends on the particular version of intuitive fit that we accept. As we discussed above, the TD approach holds that the only initially credible intuitions can be found at the high level. Accordingly, a proponent of TD must maintain that the only intuitions suited to figure as provisional fixed points lie at the high level. Theorists who accept RE believe that there may be provisional fixed points at both the high and low level while those who support BU think that they can be found only at the low level. So, in principle, provisional fixed points might be found anywhere, depending on the favoured interpretation of intuitive fit. That is, PFPA can be combined with TD, RE, and BU. Recall the above, however. We made a case against TD. If this case is accepted, we can assume that there are provisional fixed points to be found at the low level, too, as RE and BU purport. In fact, our case against consequentialism will turn entirely on low-level provisional fixed points.

Before we proceed, allow me a brief note of clarification. Some may object to our commitment to PFPA because we premised it on a particular moral-epistemological position. To explain, it is common in general epistemology as well as moral epistemology to distinguish between foundationalist and coherentist approaches to justification. And it may be alleged that our approach falls on the wrong side of this distinction. Such criticism, I think, would be misjudged. Though PFPA, as we have so far characterized it, does fall on the coherentist side, nothing that we will say below depends on an endorsement of coherentism since minor adjustments would allow us to transform PFPA into a foundationalist procedure. Let me explain.

First up, what is the essential difference between foundationalism and coherentism? The former view, I take it, assumes that all of our convictions are justified to the extent that they are either self-justifying or derivable from a self-justifying belief.[52] In contrast, the latter view is based on the idea that justification is a matter of mutual support. We cannot have self-justifying and irreversible beliefs. Rather, our beliefs are justified if and only if they fit into a web of convictions which possesses the highest possible degree of credibility *overall*. Note that PFPA does not assume that any of the intuitive convictions we use are irreversible. That is why it is called the *provisional* fixed point approach. So it falls on the side of coherentism. It is not hard, however, to transform it into a foundationalist methodology. To do that, we simply have to assume that the intuitive judgements we use to test consequentialism are not *provisional* fixed points, but *properly* fixed points. Such a modified version of PFPA (a Fixed Point Approach or FPA, for short) is defended, e.g., by Judith Jarvis Thomson. She says that she accepts PFPA

[52]Foundationalists who are non-sceptics believe, in addition to that, that there are self-justifying moral beliefs. This assumption is necessary to avoid scepticism. It is easy to see why. The foundationalist criterion of justification is perfectly compatible with the sceptical view that no moral belief fulfils it because there might, after all, be no self-justifying moral beliefs.

> with this proviso: on Rawls' account of the matter, everything is provisional, everything is open to revision, whereas I am suggesting that some moral judgements are plausibly viewed as necessary truths and hence not open to revision. (Thomson 1990, 32)

With such a *proviso* in place, our moral-epistemological approach would fall on the foundationalist side. I believe that it is not necessary to engage in any debate here. Both foundationalists and coherentists can accept our case against consequentialism based on the approach outlined above, as long as they find the intuitive judgements that we use in the argument acceptable. The only difference lies in their respective interpretations of these verdicts. Foundationalists may regard them as self-justifying moral views, while coherentists will see them merely as statements that possess a high degree of initial credibility. Who is right about this issue? In the context of our present discussion, this is largely a moot question. For this reason, we can safely put it aside.

2.4 Trolley Cases

In the previous sections, we established that, on the Rawlsian Approach, moral theories are evaluated, at least partly, in terms of their fit with our moral intuitions. There are various interpretations of this evaluative criterion: TD, RE, and BU. We argued that RE and BU, which maintain that moral theories should be evaluated, at least partly, in terms of how well they fit our low-level intuitions about cases, are the most plausible interpretations of the evaluative criterion of intuitive fit. In the previous section, then, we showed how intuitive fit can be translated into a workable, methodic approach, viz. PFPA. PFPA instructs us to look towards cases that elicit strong intuitive convictions, such that it seems reasonable to suppose that any moral theory which contradicts these intuitions seems faulty. At this point, then, it remains to be explained which kinds of cases we will use to set up our case against consequentialism and why.

There are, broadly speaking, two possibilities. We could use realistic cases – cases, that is, which have occurred in real life or are at least quite likely actually to happen. The second option is to use hypothetical scenarios which are very unlikely ever to arise in practice. As many philosophers before, we will opt for the latter. That is, we will use counterfactual and unrealistic cases which commonly go by the name "trolley cases." In what follows, we shall make some preliminary remarks about this particular sort of case. First of all, we will talk about their distinctive characteristics. The most natural way to introduce them is, I think, to look at a typical trolley case and to abstract the respective features from it. So that is what we will do. After that, we will consider two different uses for trolley cases in a moral-philosophical investigation. And we will point out how we will use them. Finally, we will explain why trolley cases seem to be particularly helpful, given the purpose at hand, before we address some worries that critics may raise about them.

2.4.1 Characteristics

Trolley cases involve a story about an agent facing a morally significant choice. This story usually revolves around a runaway trolley – hence the name – which is threatening to do some serious harm to some unlucky people. A typical example is the following scenario due to Judith Jarvis Thomson.

> **Edward's Case**
> Edward is the driver of a trolley, whose brakes have just failed. On the track ahead of him are five people; the banks are so steep that they will not be able to get off the track in time. The track has a spur leading off to the right, and Edward can turn the trolley onto it. Unfortunately there is one person on the right-hand track. Edward can turn the trolley, killing the one; or he can refrain from turning the trolley, killing the five. (Thomson 1976, 206)

Of course, the presence of a trolley is not what makes this case a trolley case. This particular detail serves merely to make for a colourful illustration of a choice situation which possesses certain distinctive features. It is these features rather than the particular story to which they are tied that make a case a trolley case.[53] We can discover them if we pay close attention to the details of *Edward's Case*.

The first thing to note is that the case strikes us, I presume, as a *tragic choice*. The above description does not state this explicitly. However, it is the most natural interpretation. And it is surely the one that is intended. We do not suppose, e.g., that the five men on the main track are "old and suicidal" and that "they'd gathered on the tracks to end their lives." (Appiah 2008, 97) We assume, quite naturally, that each person's life is valuable. And we recognize that there is no way for Edward to avoid ending at least one of these valuable lives. This is what makes his choice tragic.[54]

The second important aspect of the case is that Edward only has two options for acting. He can either do nothing or turn the trolley to the right and onto the spur.

The third feature worth highlighting about *Edward's Case* is the assumption, albeit implicit, that all normative factors that the description does not explicitly mention are absent. Further information, particularly information about the six people on the tracks, might conceivably make a difference in this situation. We assume, however, that there is no such information. For clarity's sake, let us specify what this means. The only facts that matter from the moral point of view in *Edward's Case* are the (relevant descriptions of the) acts that are available to Edward as well as their consequences. The latter, in turn, are fully described by the number of deaths that each option, respectively, will cause. All further factors that might conceivably matter are assumed to be out of the picture (cf. Wood 2011, 73–74). We

[53] We follow a terminological suggestion by Wood (2011) and Fried (2012).

[54] Note that the idea of a tragic choice is often mixed up with that of a moral dilemma. A moral dilemma is (i) a situation in which the agent is confronted with a choice between a number of options all of which are wrong or (ii) a situation in which at least two mutually exclusive options for acting are obligatory (cf. Vallentyne 1989). A tragic choice is "merely" a choice between options that are all bad.

might, e.g., suspect "that a trolley driver is a professional who is plausibly specially responsible for the trajectory of their trolley." (Mendola 2005a, 82) Alternatively, we might conjecture that one of the workers is, say, Edward's brother or friend, such that personal loyalties become relevant. We might also hypothesize that Edward is, perhaps, especially indebted to one of the workers. Though all these considerations might be morally relevant, they can be assumed to be out of the picture in *Edward's Case* because there is no explicit mention of them. I should stress that this holds, in particular, for any historical factors that might play a role. It may matter, e.g., that "one or more of the six potential victims is at fault for the coming about of the situation they now face." (Thomson 2008, 361) However, we are supposed not to make any such assumption about the history of the case.

The fourth feature about *Edward's Case* that is worth stressing is the assumption that the agent's act uniquely determines the outcome. We stipulate that, if Edward does nothing, five workers will get killed. If he steers the trolley to the right and onto the sidetrack, one person dies. There are no contingencies.

Finally, there is a fifth implicit characteristic.[55] It concerns the epistemic situation of the agent. Edward is supposed to know all of the empirical facts that the description of the case mentions. That is, he is expected to know all of his options for acting and all of their consequences.

In the remainder, we shall assume that trolley cases generally have the characteristics we just highlighted in *Edward's Case*. In brief, they can be stated thus:

Characteristic 1 (Tragic Choice)
The agent faces a tragic choice. No matter what she does, at least one person will suffer severe harm (commonly death).
Characteristic 2 (Limited Options)
The agent has a definite, limited range of options for acting.
Characteristic 3 (Absence of Normative Factors)
There are no morally relevant facts except for those explicitly mentioned. These are the options available to the agent and their respective consequences (e.g. the number of deaths).
Characteristic 4 (Determinism)
What the agent does uniquely determines the outcome of the case.
Characteristic 5 (Omniscience)
The agent knows all facts that the description of the case states.

As we noted above, trolley cases are different from real-life cases in that they are inherently unrealistic. Obviously, this has to do with the above assumptions. It may be worth noting, however, that there is a difference in kind between them (cf. Shue 2006, 231). Characteristic 1 is merely an assumption about the nature of the case. A trolley case is always tragic. This, one may say, makes it somewhat unrealistic in the sense that most cases we confront in ordinary life are not that way. Characteristics 2, 3, 4, and 5 also make trolley cases unrealistic, but in a different sense. These assumptions are never satisfied in a real case. Characteristics 2, 3, and

[55]This feature of trolley cases is discussed, however, in an exchange between Gert (1993) and Thomson (1993). See, also, Fried (2012, 2), Rosebury (1995, 499), and Wood (2011, 70) who state it explicitly.

4 are *abstractions* from the intricacies and complexities of real life. By stipulating that trolley cases possess these features, we assume away, as it were, certain morally relevant aspects of ordinary cases (cf. Gigerenzer 2008, 11; Wood 2011, 69). We assume that the agent has only very few options for acting, although it is clear that moral agents always have many options. We assume that facts which typically matter are out of the picture, although it is plain that in real-life there would always be many considerations that might be morally significant. Moreover, we assume that the agent's choice uniquely determines the outcome of the case, although there are always many factors we have to take into account in real life. Characteristic 5 is an *idealization*. We stipulate that the agent knows all relevant facts about the case, although it is clear that real-life actors never possess all the relevant information.

2.4.2 Uses

With the characteristics of trolley cases in mind, let us briefly consider two different uses for trolley cases in a moral-philosophical investigation.[56] To this end, it is useful to introduce a further trolley case. Consider the following scenario that should sound familiar.[57]

> **George's Case**
> George is on a footbridge over the trolley tracks. He knows trolleys, and can see that the one approaching the bridge is out of control. On the track back of the bridge there are five people; the banks are so steep that they will not be able to get off the track in time. George knows that the only way to stop an out-of-control trolley is to drop a very heavy weight into its path. But the only available, sufficiently heavy weight is a fat man, also watching the trolley from the footbridge. George can shove the fat man onto the track in the path of the trolley, killing the fat man; or he can refrain from doing this, letting the five die. (Thomson 1976, 207–208)

Edward's Case and *George's Case* are similar. In each case, the respective agent has two options for acting. And in each case, one of these options leads to the death of five people, while the other leads to the death of only one person. I assume, however, that our intuitions as to the permissibility of the agent's choices differ between the two cases. At any rate, most people believe that it is at least morally permissible for Edward to kill one instead of five. A majority, however, feels that it is impermissible for George to kill the fat man by pushing him off the bridge.[58] Trolleyologists (as philosophers who deal in trolley problems are sometimes called) have commonly used these facts about our intuitions regarding these cases "for

[56] The distinction we use corresponds to Karl Popper's distinction between the different uses of thought experiments in science (and especially in quantum theory), viz. the *apologetic* use and the *critical* use (cf. Popper 1959/2005, 464–480).

[57] We have already come across this scenario on page 22.

[58] These conjectures were made by Thomson (1985). In the meantime, a lot of empirical evidence has been piled up to support them. Important sources can be found in Greene (2008, 42).

the purpose of unearthing principles of permissible harm." (Kamm 2007, 4)[59] The idea is to go through a series of such cases, to consider our intuitive responses to find provisional fixed points, and to formulate moral principles and theories which capture these fixed points. The exercise is analogous to that of an empirical scientist fitting a curve to her data points.

This, however, is not the only possible use of trolley cases. Philosophers also employ them with critical intent, that is, to test moral theories (cf., e.g., Tännsjö 2011, 295). Here, the idea is to look at a given theory and to check whether it matches provisional fixed points in a specific case or series of cases. We may, e.g., test theories regarding their implications in *Edward's Case* and *George's Case*. That is, we may reject all doctrines that do not imply that Edward should steer the trolley to the right, killing the one. And we may reject all doctrines that do imply that George should push the fat man, saving the five. The analogue to this second use of trolleyology is the case of the empirical scientist who critically tests a theory by examining whether it does, in fact, capture all available data points.[60]

The distinction between these two uses is important because some worries about trolley cases appear to relate solely to the first one. It should be noted, then, that we will use trolley cases only in the latter way. That is, we will use them only to test consequentialism. We will not argue for an alternative moral theory.

2.4.3 Pros and Cons

With the various features and uses of trolley cases in plain view, we can address some of the pros and cons of trolleyology, starting with the pro side. There are two main reasons for using trolley cases rather than more realistic scenarios. These relate to their aforementioned features. One is specific to our investigation. The other is more general.

The first reason lies in the desire to reduce complexity. As will get clear below, one of the difficulties about any study of consequentialism is the fact that there are so many versions of it. Given the scope of this inquiry, going through all of them is an unmanageable task. Trolley cases, however, allow us shortcuts. By using them, it is possible to set aside certain varieties of consequentialism *ab ovo*. Here is why. The differences between the various consequentialist doctrines manifest only

[59] See, also, Wood (2011, 67).

[60] Of course, the scientist need not immediately reject the theory if it turns out that it does not capture all data points. There are always ways of accounting for recalcitrant evidence which are compatible with the truth of the theory (cf. Lakatos 1970). Similarly, a moral theorist need not reject a moral principle if it violates one or more out of a number of provisional fixed points. As we explained on page 40, PFPA is subject to a *proviso*. Should it turn out that no moral theory can, in fact, accommodate all of our provisional fixed points, the fact that a given theory violates one of these points cannot, by itself, count as counter-evidence against it.

in particular kinds of cases. Depending on the case at issue, it may, therefore, be unmotivated to distinguish between certain varieties of consequentialism.

An example should help to drive home the point. It is common, e.g., to differentiate between *subjective* and *objective* versions of consequentialism (cf., e.g., Howard-Snyder 1997). The distinction is, roughly, this. On Subjective Consequentialism, the moral status of an act is determined by the consequences that the agent expects. In contrast, Objective Consequentialism turns on the actual result of the agent's choice. To resolve whether her act is right or wrong, it looks towards objective consequences. Note, then, that the difference between these two versions of consequentialism is only relevant in cases where the agent's epistemic situation is imperfect. We have to assume that she cannot know for sure what the consequences of her act will be so that subjective and objective results can, in fact, come apart. If, on the other hand, the agent knows for sure what will happen if she chooses this or that act, subjective and objective consequences will coincide and so will the moral verdicts of Subjective and Objective Consequentialism. Now, trolley cases make, as we know, an idealized assumption about the agent's epistemic situation. She is supposed to have perfect knowledge of all morally relevant facts of the case (Characteristic 5). This, of course, includes the objective consequences of her options. Since she knows this, subjective and objective outcomes coincide, and so do the verdicts of Subjective and Objective Consequentialism. This *must* be the case! Hence, it eliminates the motivation for distinguishing between these two variants of consequentialism. This is, of course, only an example. As we will see in more detail below, trolley cases do not only take away the motivation for a distinction between Subjective and Objective Consequentialism. They also make superfluous the difference between Direct and Indirect Consequentialism and between consequentialist doctrines that subscribe to different theories of individual well-being (e.g. Welfarism Hedonism, Welfare Preferentism, and so on). This will help to reduce the workload considerably.

The second, more general reason why trolley cases seem useful is that they allow us to clarify our intuitions. This has to do with their simple make-up and their comparatively little complexity. Realistic cases, in contrast, can be very fuzzy. There are many options, a lot of normative factors to consider, and other relevant aspects that do not play a role in trolley cases. Under these complexities, our intuitions may give out (cf. Nagel 1986, 180). Moreover, even if we do have an intuition about a case, it is unclear whether it is reliable. For it may be, as we observed previously, that we fall prey to moral "blind spots." That is, we may end up paying too much attention to certain factors, while ignoring others. This, it seems, is not as likely to happen in a trolley case. Here, the relevant facts are reduced to a minimum such that we may assume that anyone can handle the cognitive load.[61]

[61] A further consideration that might be brought up to motivate the use of trolley cases is given by Amartya Sen. He writes that "in many of the common cases, intuitions based on quite different principles tend to run in the same direction, so that it is impossible to be sure of the basis of an overall judgment." (Sen 1982, 14) Therefore, it may be hard for a moral theorist to establish that her favoured theory is the best explanation for our moral intuitions if only common cases are used.

So much for the plus side. It is evident, however, that the advantages of the methodic use of trolley cases (trolleyology, henceforth) come at a cost. Trolley cases have a very simple, idiosyncratic makeup. As a consequence, certain types of ethical problems cannot be addressed in a trolleyological investigation. Take, e.g., the ethics of risk and uncertainty. It is undoubtedly an important theme in moral theory that consequentialists have had interesting things to say about (e.g. Norcross 1998). Now, as we discussed above, trolley cases assume that the decision of the agent necessitates a given outcome (Characteristic 4). Hence, they cannot be used to address moral issues that may arise in the context of risk and uncertainty. By using trolley cases, we will, therefore, inevitably miss important aspects of the moral-philosophical debate that pertain to these phenomena. Philosophers who are especially interested in discussing them may, therefore, regard trolleyology as a flawed method. I believe, however, that it is possible to address their reservations. To do this, we should remind them that we use trolley cases with a specific purpose in mind. We seek to construct an argument that makes plausible the claim that all forms of consequentialism should be rejected. To do this, we do not have to address all ethical issues on which consequentialism may have something to say. All we need to do is to demonstrate that there is at least one serious objection to all forms of consequentialism. If we succeed in doing this by using trolley cases, we may skip over many interesting philosophical questions. However, we will, nevertheless, accomplish what we set out to do.

This said, we should turn to some objections that seem to be more fundamental and more severe. Before we do that, however, allow me to express my discontent with the current state of the debate. The use of hypothetical cases in ethics and trolley cases, in particular, has "become so common that many philosophers hardly notice it and if they do, find it unproblematic." (Elster 2011, 241–242) In fact, the principal exponents of trolleyology usually do not bother to justify their methodology properly. What they say about it hardly ever surpasses the stage of mere explanation, even though objections to trolleyological thinking have been piling up for years. This is a lamentable fact. A systematic and comprehensive investigation of the virtues and limitations of trolleyology is surely in order. However, given the limited scope of this inquiry, it is not a task we can take on here. Nevertheless, we shall try, at least, to make our trolleyological method plausible.

This having said, let us turn to the objections to trolleyology. First up, we should demarcate two sorts of scepticism about it. One kind of worry has to do with the fact that the methodology relies on our low-level intuitions about cases. In Sect. 2.2, we addressed this concern at some length. We concluded that, to the extent that we can trust our intuitions at all, we do not seem to have any reason to distrust low-level intuitions in particular. Hence, we shall set this particular worry aside.

If our intuitions about these cases can be explained by a large number of normative factors, the moral theorist will have a hard time arguing that *her* explanation should be chosen. "In order to do the discrimination," Sen says, "we choose examples such that different principles (. . .) push us in different directions." (Sen 1982, 14) Sen's reasoning, I believe, provides a good motivation for the constructive use of trolley cases, but is less relevant to our destructive use.

Instead, we shall focus on objections that philosophers voice who are otherwise sympathetic to the idea that our intuitions regarding cases can be valid but insist that the particular features of trolley cases disqualify them for the purpose of moral inquiry. These objections mark, as it were, a "family quarrel" (Elster 2011, 242) between philosophers with similar epistemological inclinations (either towards BU or RE).

Objection 1
People disagree about trolley cases

The first objection that we shall address relates to one of the points that we made in our general discussion about the reliability of intuitions.[62] Recall the quote by Henry Sidgwick that we came across above. Sidgwick says that "if I find any of my judgments, intuitive or inferential, in direct conflict with a judgment of some other mind, there must be error somewhere." (Sidgwick 1907, 342) This, in turn, should reduce the confidence in my judgement. Obviously, the same goes for intuitions about trolley cases. If there is genuine disagreement about them, we should be cautious. Now, it may be suggested that the empirics of trolleyology show that people do disagree about trolley cases. This may give rise to something like the following argument.

(P1) If people disagree in their intuitive judgements about a case, this makes everyone's intuitions about that case initially incredible.
(P2) People disagree in their intuitive judgements about trolley cases.
(C1) Everybody's intuitions about trolley cases are initially incredible. (from P1, P2)
(C2) Trolleyology is an invalid method. (from C1)

As it stands, it is unclear what this argument says. For one thing, it contains a concealed quantification – assuming, of course, that it is formally valid. C2 follows from C1 only if we interpret C1 as a universal statement. Trolleyology is an invalid method only if *all* our intuitions about trolley cases are unreliable. Because then it would be impossible to find any trolley case to which it could justifiably be applied. If, however, we can have initially credible intuitions about trolley cases and if we restricted the application of the trolleyological method to these cases, then it would seem to be unobjectionable. Hence, we have to assume that C1 is a statement about *all* our intuitions. Furthermore, this version of C1 follows from P1 and P2 only if we also interpret P2 as a universal assertion. Plainly, if we interpret P2 merely as saying that people have different intuitions about *many* trolley cases, it would not follow that *all* intuitions about trolley cases are initially incredible. In that case, there might be some intuitions that are in fact reliable. And the success of a trolleyological investigation may largely be seen as a matter of finding them. So P2 and C1 have to be interpreted as *universal* statements. This, in turn, means that objectors have to interpret P2 in its strongest and least plausible form.

[62] I am indebted to Michael von Grundherr for making me aware of this objection.

The precise purport of P1 and P2 is also unclear. Both employ the notion of interpersonal disagreement. It is important to spell out what this involves. To this end, let us look at an example. Suppose that we present 1000 people with a case. 999, say, share an intuition about it. One person, however, reports a different intuition. Is it adequate, then, to say that people disagree in their intuitive judgements about this case? I believe that this is not so in any relevant sense. We are interested in cases of disagreement that would cast doubts on our intuitions. Their reliability, given a certain level of disagreement, depends on a number of factors besides the fact of disagreement itself. This is easy to see. Even on the assumption that our intuition about a given case is reliable, we would still expect to find some disagreement on it. When we ask people what they think about this or that case, we would expect some people to misapprehend it. Moreover, we would expect some people to interpolate additional assumptions that are not intended (contrary to Characteristic 3, mind you). And we may even suspect that some people are merely joking about their answer. We have no reason, then, to distrust our intuitions about trolley cases if it is possible to explain the level of disagreement among them by factors such as these. Hence, the notion of disagreement that is relevant here is *substantial* disagreement. P1 should, hence, be read as saying that everybody's intuitions about a case are unreliable if people disagree substantially in their intuitive judgements about that case. And P2 should be interpreted as saying that people disagree substantially about all trolley cases.[63]

So much, then, for the interpretation of the argument. Let us consider now whether the premises, P1 and P2, are plausible. P1 certainly is. Following Sidgwick, it makes sense to take our intuitions with a grain of salt if we find that they are subject to substantial disagreement. P2, however, appears to be false as a matter of empirical fact. It does not seem to be true, that is, that people substantially disagree on *all* trolley cases. Rather, there are some on which they agree and some on which they do not agree. Hauser et al. (2007), e.g., conducted a study of the moral intuitions of over 5000 subjects from 120 countries. They found that in the so-called "loop case" (Thomson 1985, 1402) 56 % of the people asked judged a given act permissible, while 44 % opposed this view.[64] However, Hauser et al. (2007) report a greater measure of agreement, ranging from 72 % to 88 %, in three other cases. These included one scenario that resembles Philippa Foot's original trolley case (*Edward's Case*) and one situation that is fashioned after Judith Jarvis Thomson's Fat Man Case (*George's Case*). This gives us reason to suspect that it

[63] It is hard to pin down, of course, what a reasonable threshold for substantial disagreement is.

[64] The description of the case was as follows: "Ned is walking near the train tracks when he notices a train approaching out of control. Up ahead on the track are 5 people. Ned is standing next to a switch, which he can throw to turn the train onto a side track. If the train hits the object, the object will slow the train down, giving the men time to escape. The heavy object is 1 man, standing on the side track. Ned can throw the switch, preventing the train from killing the 5 people, but killing the 1 man. Or he can refrain from doing this, letting the 5 die." (Hauser et al. 2007, 5) The question that was asked was "Is it morally permissible for Ned to throw the switch?" (Hauser et al. 2007, 5)

may be possible to find cases about which people do not disagree substantially and may suggest that we need not regard our intuitions about these cases as initially incredible.

Though we should, I believe, reject the argument from disagreement against trolleyology, it highlights a methodological point of some importance: When we construct arguments based on our intuitions about trolley cases, we had better check whether there is a substantial disagreement among them. In recent times, some of the foremost trolleyologists have neglected this practice and objections to *their* (ab)use of trolley cases may be raised quite fairly. Frances Kamm, e.g., does not seem to worry at all whether her intuitive judgements are agreeable to others, even if they figure as crucial premises in her argument.[65] It is no surprise, therefore, that other philosophers disagree with her to the extent that they see no common ground with her *at all*. Some even report that they "have found *no one* who agrees with her." (Norcross 2008, 66; emphasis in the original, NM) This sort of embarrassment is one that we shall seek to avoid in our case against consequentialism. Of course, since this is not an empirical study, we have no way of knowing for sure whether there is, in fact, a substantial disagreement on the intuitive judgements that we use to make our case. But let us, at least, be open to that kind of empirical refutation. And let us try to use only intuitive judgements that we may reasonably take to be entirely uncontroversial.

Objection 2
Trolley cases allow no general moral conclusions.

A further common objection to trolley cases consists in saying that they provide an inadequate basis for generalizations. It may be argued that it is illegitimate to draw any substantive lessons from trolley cases because many of the principles that philosophers have derived from them "do not produce the 'right' answer if applied beyond trolley cases." (Fried 2012, 13) At best, one might insist, trolleyologists may conclude that a given principle holds in a particular trolley case. However, a more general conclusion, it may be claimed, is unfounded.

This invective concerns, I think, not trolley cases per se. It concerns only one of the two uses of trolley cases. To be more precise, it concerns only the way in which we will *not* employ them. Here is why. Recall the distinction that we made above between the two primary purposes of trolley cases. We can use them in a constructive way, and we can employ them in a critical or destructive way. The idea behind the former use is to look at a series of trolley cases and to induce from them a general moral principle (cf. Kamm 2007, 4). The other is to use them as critical tests for given moral theories (cf. Tännsjö 2011, 295). As Popper (1959/2005) pointed out, there is, from a logical point of view, an asymmetry between these two enterprises. When moral theorists construct moral principles based on trolley

[65] In fact, much unlike Sidgwick, Kamm explicitly advises her readers to ignore the intuitive judgements of others. She does that since she believes that "much more is accomplished when one person considers her judgments and then tries to analyze and justify their grounds than if we do mere surveys" (Kamm 2007, 5).

cases, they do so, usually, hoping that these principles will match our provisional fixed points not just in those cases, but in *all* cases.[66] However, all they can say for sure is that their conjectured principles match the provisional fixed points in the particular trolley case(s) they have looked at. This is analogous to the case of the empirical scientist who cannot be sure that the curve she has plotted based on given data can accommodate the next data point. It is always possible that the very next trolley case shakes the firm confidence theorists put in their principles. And it is possible, furthermore, that their principles yield an entirely wrong conclusion when applied to real-life cases.

This, I believe, is the point of the objection. There is certainly something to it. It may, in fact, be very problematic to use trolley cases to derive moral principles and to then generalize them. Maybe real-life cases have important features – e.g. risk and uncertainty – which call for moral principles that are entirely different from those which suggest themselves in trolley cases. Note, however, that our critical use of trolley cases is unaffected by this criticism. Here, the aim is not to derive moral theories which match certain fixed points. Rather, the idea is to test specific doctrines and to check whether a case exists in which they give an answer that appears plainly wrong. Once we have established that a given theory does, indeed, give such a highly problematic answer, this is a *fact* we can work with. And it is a fact that does not change. As it turns out, then, the objection does not apply to our use of trolley cases.

Objection 3
Trolley cases suppose that normative factors are additively separable.

Another reason to reject trolleyology is to say that it employs a strategy that "relies on an underlying assumption concerning the role of [normative; NM] factors – an assumption that is questionable and should probably be rejected."[67] (Kagan 1988, 12) To explain, trolleyological inquiries do not, for the most part, rely only on one case. They rely on *pairs* or *series* of cases that are sometimes used to construct what Shelly Kagan calls "contrast arguments." The idea is that we take one case, vary one factor, holding everything else fixed, and compare the contrast case that results from this modification to the original case. If we find that the moral evaluations of the two cases differ, we conclude that this is due to the varied factor. If we find that nothing changes, we conclude that the factor does not play a role. In and of itself, this procedure is not objectionable. But here comes the kicker. Since the conclusion that a factor matters (or does not matter) in a particular case is quite unexciting, we shoot for a bolder claim and generalize our finding, employing what Kagan calls the "ubiquity thesis." The idea is that "if variation in a given factor makes a difference *anywhere*, it makes a difference *everywhere*." (Kagan 1988, 12; emphasis in the original) Hence, if a factor contributes to the normative evaluation of a given case, we conclude that it *always* makes this contribution. Moreover, if

[66] As we shall see below, however, some theorists deny this.

[67] This problem is also acknowledged and responded to by Kamm (2007, 345–367; esp. 348–349). See, also, Kamm (1983).

it does not play a role in that one instance, we take this to indicate that it *never* plays a role. The reason we believe the ubiquity thesis, Kagan argues, is that we think of normative factors as having additively separable weight. That is, we picture their weights like numbers in an addition equation. Each number on the left-hand side of the equation increases the value of the number of the right-hand side by a certain amount. And it does that independently of the values of the other summands. E.g., adding 5 to a sum of numbers always increases the value of that sum by 5. Analogously, normative factors that contribute to the rightness of a given act in one case are thought to make the same contribution to the rightness of acts in other cases.

Now, what is the problem with all of this? The problem is that this way of thinking about normative factors seems to be flat out incompatible with many respectable moral views. Many of us are *holists* about factors. To illustrate, most of us would agree that the fact that an act alleviates suffering is a morally relevant factor that generally counts in its favour. At the same time, however, many of us might believe that there is no reason to do a particular act, even though it alleviates suffering. E.g., when a guilty person is punished, there is perhaps no reason to bring relief to that person because she *deserves* to suffer.[68] This, at any rate, is what many people are inclined to think. Holding such a belief system, however, is inconsistent with the ubiquity thesis. According to the ubiquity thesis, if the fact that the act alleviates suffering counts in its favour once, it always has this effect.[69]

In regards to Fried's objection, we said that it applies only to constructive trolleyological arguments, while our inquiry seeks to establish a critical conclusion. The additive assumption and the ubiquity thesis, however, seem to underlie both the constructive and the critical use of trolleyology. Advocates of moral theories imagine trolley cases which support the significance of the factors which, according to their theory, are important. Critics of moral theories conjure up cases in which these factors seem to be irrelevant. Both generalize their findings using the ubiquity thesis. Advocates conclude that the respective factors are always relevant, while critics infer that they are always irrelevant. Hence, the reply that we gave to Fried's objection will not do here. Instead, we should draw attention to the fact that it is not always necessary for critics of a moral theory to show that the factors that the theory *always* takes to be relevant are, in fact, *never* relevant. It may be enough to show that a theory violates certain provisional fixed points in one instance. In fact, this is precisely what we shall attempt to do in our case against consequentialism. At no

[68]This position is called *retributivism* and is commonly associated with Immanuel Kant, who expressed the view in his *Metaphysics of Morals* (*Die Metaphysik der Sitten*). See, in particular, his infamous thought experiment of the dissolving civil society (cf. Kant 1803, 229).

[69]Note, however, that the chosen illustration is not a conclusive demonstration, as it presupposes a particular model of normative factors on which the property of an act to alleviate suffering is seen as an autonomous factor. This model can be rejected. Alternatively, we may distinguish between acts that alleviate the suffering of an innocent person and acts that alleviate the suffering of a guilty person that results from a just punishment. We can, then, take the former to be a right-making feature and the latter to be a wrong-making feature. This would resolve the difficulty in the present case.

point throughout the inquiry shall we generalize our conclusions beyond the level of the individual case. Hence, the ubiquity thesis has no role to play in our argument. We can allow ourselves to remain agnostic about it. [70]

Objection 4
Trolley cases are outlandish.

A further objection to trolleyology is to say that trolley cases are outlandish (cf., e.g., Kagan 1998, 76–77). In and of itself, this does not seem to be a problem. So what does the objection consist in precisely?

One interpretation is this. Since the scenarios that trolley cases present are so outlandish and unlikely, there is no reason to suppose that we have reliable intuitions about them. We may feel that our immediate judgement is robust. But this is a mistake. Our moral intuitions are not fit to judge cases of that kind. They evolved to help us deal with "normal" scenarios that we are likely to encounter on a daily basis. Hence, we should not trust them in freakish and unusual cases, such as trolley cases (cf. Singer 2005).[71]

This variant of the objection is highly implausible, as Allen Wood points out. "It is extremely rare," he says, "for a man to lure teenage boys into his apartment, then kill, dismember and eat them (. . .). But the rarity of such cases does not lead us to mistrust our moral intuitions about these cases" (Wood 2011, 69).

Another interpretation of the objection is this. Since trolley cases are unlikely ever to arise in practice, it is not fair to use them as tests for moral theories which aim to assist us in making *practical* choices. Those who bring up this objection seem to underestimate the tremendous ambitions that consequentialists have commonly had. They aspire to offer us a *universal* standard of right and wrong which applies, as Jeremy Bentham zealously professed, to "every action *whatsoever*" (Bentham 1838, 1; emphasis added, NM). Therefore, they seem to be in no position to cry foul when their critics invoke cases that are unlikely ever to arise in practice. Given consequentialists' "universal pretensions," their theories are, as Robert Goodin has emphasized, "absolutely fair game for purveyors of such fantasies" (Goodin 1995, 6).

There is an obvious objection that an objector may give to this reply. She can say that we should, perhaps, drop the "universal pretensions" of our moral theories and understand ethics, for once, as a *practical* discipline. Accordingly, we should eschew hypothetical examples and should use realistic scenarios to test doctrines. Or, as Thomas Pogge says, "[w]hat does it matter that our morality is inapplicable to the life context of fictitious Martians or of the ancient Egyptians, so long as it

[70] However, see Sorensen (1998, 272–273) for a critical rejoinder to Kagan's argument.

[71] Hare (1981, Chap. 8) gives a similar justification for Objection 4. As he argues, our intuitions are the product of our moral upbringing. He believes that "however good these may have been, they were designed to prepare [us] to deal with moral situations which are likely to be encountered" and that, therefore, "there is no guarantee at all that they will be appropriate to unusual cases." (Hare 1981, 132).

provides reasonable solutions to our problems."[72] (Pogge 1990, 660) On this view, the testing of theories against surreal scenarios is useless at best, as these situations are irrelevant in practice. Moreover, it may even be positively harmful because the use of hypothetical cases may lead us to reject moral theories that give entirely satisfactory answers to the practical problems they are intended for.[73]

This plea may be a fair point. It is noteworthy, though, that not all moral theorists are in a dialectical position to make it. Thomas Pogge can consistently raise it because he believes that the endorsement of moral principles "is consistent with their limited range."[74] (Pogge 2000, 138) As a pluralist about moral realms, he believes that principles vary across domains, where the domain of *real* or *possible* cases may be one to which specific principles apply – principles that do not apply to *outlandish* ones.[75] Consequentialists, on the other hand, are *monists* in the sense that they claim that there is precisely one moral criterion which applies to all acts and under all circumstances. Hence, they cannot put forward such a reply. In doing so, they would *ipso facto* abandon their moral theory.

Let me state, then, by way of conclusion, that the use of trolley cases is controversial. However, it appears to be rather unobjectionable, given the purpose to which we will put these cases in our subsequent investigation.

2.5 Summary

Let us sum up. The aim of our inquiry is to reject all versions of consequentialism. To develop an argument to this effect, we need to understand how moral theories can be evaluated and criticized. In this chapter, we tried to do just that.

In Sect. 2.1, we investigated the Rawlsian Approach, which seems to be the *modus operandi* in moral philosophy these days. It says, roughly, that a moral theory is acceptable to the extent that it fits our moral intuitions, is consistent, and establishes explanatory connections. As we discussed, this idea can be factorized into two sub-criteria, viz. *intuitive fit* and *coherence* which can, in turn, be factorized into two further criteria, viz. *consistency* and *systematicity* (or *connectedness*). In the debate about consequentialism, intuition-based arguments occupy center stage. Thus, we decided to base our argument on the criterion of intuitive fit.

[72] Similar views can be found, e.g., in Rawls (1951, 182 and 2003, 71), Hare (1981, 47–48), and Miller (2008, 44).

[73] An argument much like that was suggested to me by Andreas Suchanek in personal conversation. I believe that this way of thinking is common amongst scholars whose predominant focus is applied ethics.

[74] See, in particular, sections VIII through XIII in Pogge (2000).

[75] The sense in which the term "pluralist" is used here should not be confused with the sense in which it was used above. In Sect. 1.2.1, we called a moral theory pluralist if it contained more than one foundational moral principle. Here we call it pluralist if it contains different moral principles for different realms. A moral theory can be pluralist in the one sense but not in the other.

In Sect. 2.2, we talked about three interpretations of intuitive fit, viz. the Bottom-Up Approach (BU), the Reflective Equilibrium Approach (RE), and the Top-Down Approach (TD). This differentiation is based on a distinction between two types of moral intuitions, viz. high-level intuitions that concern abstract and principled questions and low-level intuitions which relate to concrete cases. BU is the view that only low-level intuitions are initially credible and that one should evaluate a moral theory according to its fit with them. RE is the more ecumenical view that both high-level and low-level intuitions can be initially credible and that a moral theory should, therefore, be evaluated in light of its overall fit with both of them. TD is the view that only high-level intuitions are initially credible and that we should judge a moral theory according to its fit with them. We argued that TD should be rejected and that either BU or RE is the correct view. This is important because our argument in Chap. 5 will rely on the assumption that intuitions about moral cases are admissible in moral inquiry.

In Sect. 2.3, we then considered how we can develop a workable methodic procedure for our investigation based on the evaluative criterion of intuitive fit. This step was necessary because intuitive fit does not, in and of itself, provide a testing procedure for moral doctrines. It merely gives us a philosophical ideal, viz. that our moral theories should fit our moral intuitions. We introduced and discussed the Provisional Fixed Point Approach (PFPA). The idea behind it is this. To test theories, we check them against provisional fixed points in our thinking. These provisional fixed points are intuitive convictions which are so strong that it seems reasonable to expect that an acceptable moral theory should be able to match them. If it does not, we can justifiably reject it. This conclusion is, of course, provisional in nature. It may turn out that no moral theory can fit all our provisional fixed points. In that case, the conclusion may not hold. Whether that is, in fact, the case is, however, a question we cannot address, given the limited scope of our inquiry. Having explained the basic idea of PFPA, we tried to make it more concrete by linking it with some of the points we had made in the second section of this chapter. We noted that we could, in fact, combine PFPA with BU, RE, and TD. TD, which we had rejected, would rule out provisional fixed points at the low level. However, BU and RE, which we did not exclude, allow them. Hence, we concluded, that PFPA in conjunction with either BU or RE permits us to draw on our intuitions about cases. This, in fact, is how we will proceed in our argument in Chap. 5.

In Sect. 2.4, we noted that, though PFPA allows us to use cases, it does not give us any guidance as to the kinds of cases we should use. We looked at trolley cases and concluded that they are suitable for the task ahead. We started by looking at their characteristics. Then, we went into their possible uses. Finally, we considered some objections to them that critics raised in recent times. Our main point was that valid criticisms of the methodical use of trolley cases do not seem to concern the way in which we will use them. They are directed only at the constructive use, while we are interested in employing them with a critical intent only.

http://www.springer.com/978-3-319-39248-6

The Case Against Consequentialism Reconsidered
Mukerji, N.
2016, XXIII, 245 p. 2 illus., Hardcover
ISBN: 978-3-319-39248-6