



Intuitions are never used as evidence in ethics

Tomasz Herok¹

Received: 13 November 2021 / Accepted: 23 December 2022 / Published online: 24 January 2023
© The Author(s) 2023

Abstract

One can often hear that intuitions are standardly “appealed to”, “relied on”, “accounted for”, or “used as evidence” in ethics. How should we interpret these claims? I argue that the typical understanding is what Bernard Molyneux calls “descriptive evidentialism”: the idea that intuition-states are treated as evidence of their propositional contents in the context of justification. I then argue that descriptive evidentialism is false- on any account of what intuitions are. That said, I admit that ethicists frequently rely on intuitions to clarify, persuade, discover, or to support things other than the intuitions’ contents. The contents of intuitions are also commonly used as starting premises of philosophical arguments. However claims about these practices need to be sharply distinguished from the prevalent dogma.

Keywords Intuitions · Evidence · Metaphilosophy · Philosophical methodology · Reflective equilibrium · Method of cases

1 The dogma

The view that moral philosophy somehow depends on intuition has become a prominent part of the profession’s self-image in recent decades. Here are some examples of how it is expressed:

The most common method in normative ethics is piecemeal appeal to intuition. ‘It follows from what you say that it would be all right to do such and such, but that’s counter-intuitive, so you’re wrong.’ (Griffin, 1988, p. 1)

The appeal to intuitions is a pervasive strategy in contemporary philosophical discourse. A good philosophical theory is widely taken to be one that gives an adequate account of our intuitions. Ethical theory is no exception. (Audi, 1993, p. 295)

✉ Tomasz Herok
t.herok@lancaster.ac.uk; tomasz.herok@gmail.com

¹ Department of Politics, Philosophy and Religion, Lancaster University, Bailrigg, Lancaster LA1 4YW, UK

Anyone who reflects on the way we go about arguing for or against moral claims is likely to be struck by the central importance we give to thinking about cases. Intuitive reactions to cases—real or imagined—are carefully noted, and then appealed to as providing reason to accept (or reject) various claims. (Kagan, 2001, p. 44)

Philosophers these days frequently elicit “our intuitions” about this or that and appeal, implicitly or explicitly, to our feelings and sentiments, and to moral consensus. They invent imaginary cases and tell us bizarre stories which are intended to illuminate these intuitions. Pick up any recent journal or Moral Problems anthology, and it seems as if everyone is going about ethics in a similar way. (Shaw, 1980, p. 127)

Many moral theorists have relied on intuitions in both building up and challenging theories. (Kamm, 2007, p. 425)

Many contemporary ethicists like to treat moral intuitions as evidence, akin to experimental data that are to be explained by theories. (Thagard, 2010, p. 202)

Ethicists often appeal to moral intuitions in defending a theory. In this practice, the contents of intuitions are taken to support moral beliefs in a way that is often compared to the way the contents of perception support empirical beliefs. (Kauppinen, 2015, p. 169)

It appears that in moral reasoning, moral intuitions play the same role which observations do in science: we test general moral principles and moral theories by seeing how their consequences conform (or fail to conform) to our moral intuitions about particular cases. (Boyd, 1988, pp. 184–185)

It is hard to imagine any way to develop a moral theory without relying on moral intuitions at all. How could you choose among consequentialism, Kantianism, contractarianism, and virtue theories without appealing to moral intuitions at some point in some way? (Sinnott-Armstrong et al., 2010, p. 246)

In the end, all ethicists appeal to intuition. They can do no other. (Bedke, 2008, p. 266)

I am going to argue that all these claims, as well as countless similar ones, are false: they all refer to a non-existent practice. This is not to say that the sentence “intuitions are relied on in ethics” is false on any reading, however none of the true readings has much to do with what philosophers have in mind when they make claims like the ones I have just listed.

I believe that the myth of relying on intuitions in ethics is part of a greater myth of relying on intuitions in philosophy. Nevertheless the article focuses on ethics, especially in terms of choice of examples. This is because it is sometimes openly stated (Weatherson, 2003, p. 1; Griffin, 1988, p. 2) and sometimes implied (Brandt, 1979; p. 16, McMahan, 2013) that the way intuitions are treated in ethics is different from the way they are treated in other areas of philosophy. Doing full justice to such views would require carefully analysing examples of the alleged reliance on intuitions from a variety of philosophical disciplines and explaining why some philosophers believe ethics is methodologically unique, which would be unwieldy in an article-length text. The reader should be reminded that most of my argument applies to philosophy in general, however a stronger case against the idea of using intuitions outside ethics would require additional work.

2 State vs content

Until recently the view that philosophers rely on intuitions did not even have a name, which is not very surprising given how universally accepted it was. Currently the most common term in the literature seems to be “Centrality”, coined by Herman Cappelen. Here is how Cappelen defines it:

Centrality (of Intuitions in Contemporary Philosophy): Contemporary analytic philosophers rely on intuitions as evidence (or as a source of evidence) for philosophical theories. (Cappelen, 2012, p. 3)

What is the difference between “evidence” and “source of evidence”? Some have pointed out that “relying on an intuition that p as evidence” is ambiguous between “relying on p , which is intuitive, as evidence” and “relying on the fact that p is intuitive as evidence”. In other words, what constitutes evidence can be intuitions in the propositional content sense or intuitions in the mental state sense (Molyneux, 2014, p. 443; Deutsch, 2015, pp. 35–39). Proponents of Centrality sometimes endorse the latter, and sometimes the former—but with an addition that intuition-states then serve as a source of evidence (Cappelen, 2012, p. 13).

Both the propositional and the mental state reading can also be endorsed simultaneously. For example, George Bealer argues that there is an element of philosophical inquiry when the propositional content is used as evidence and an element when the fact of intuiting this content is used as evidence (Bealer, 1998, p. 205). What I am going to object to is the intuiting element only, as I believe that philosophers often rely on propositions which merely *happen* to be intuitive. Consider the statement: “Contemporary analytic philosophers rely on propositions formulated by carbon-based life forms as evidence for their theories”. It sounds odd, as it is *pragmatically inappropriate*: there is little point of bringing up the chemical composition of creatures who formulate propositions in the context of discussing philosophical evidence. However, pragmatics aside, the statement is not false. In my view, the statement “Contemporary analytic philosophers rely on intuitive propositions as evidence for their theories” has a similar status. It may be odd or unhelpful, but it is not false.

Cappelen’s take on this issue is somewhat different—he does not believe that philosophers rely on intuitions in the propositional content sense. This is because of his scepticism about what intuitions, as philosophers use the term, are—according to Cappelen, philosophical intuition-talk is “a linguistic practice bordering on gibberish” (2012, p. 61). I do not find his argument persuasive, but, more importantly, I do not find it *necessary*. The only reason why Bealer and others say things like “contents of intuitions count as evidence” is that they believe that the state of being intuitive also counts as evidence. As I am going to argue, when the latter is rejected, uttering the former becomes pragmatically inappropriate in most contexts, and it does not matter whether it is true, false or neither—at least not to someone interested in the problem of philosophical evidence. I do, however, agree with Cappelen that those who prefer to speak of intuition-states as *a source of evidence* (like, for example, Nado, 2017) are mistaken. And like Cappelen, for brevity’s sake I am also going to refer to what they call “a source of evidence” simply as “evidence”.

3 Own content vs other content

To be more precise, I am only going to attack one version of the mental state interpretation. Note that “relying on the fact that p is intuitive as evidence” is itself ambiguous between “relying on the fact that p is intuitive as evidence for p ” and “relying on the fact that p is intuitive as evidence for q ”. It is only the former interpretation that I am going to object to.

For example, Robert Nozick in his famous thought experiment asks whether you would plug into a machine that could give you any experience you wanted, indistinguishable from experiencing reality (Nozick, 1974, p. 42). He argues that for many people their “first impulse” is to say no, even if they might later change their mind upon reflection (Nozick, 1989, p. 105). I do not want to deny that Nozick is using an intuition as evidence against psychological hedonism (the view that all that motivates us is pleasure). Philosophers occasionally evaluate psychological claims and appealing to people’s intuitions—in the form of “it is intuitive that p , therefore q ”—may be a way to do it. Nor do I want to deny that Nozick also tries to provide evidence against *ethical* hedonism (the view that all that matters is pleasure). My quarrel is with the idea that Nozick is offering the intuition about the experience machine as evidence against ethical hedonism. On this reading the intuition is used as evidence *for its content*: one should not plug oneself into the machine because it is intuitive that one should not plug oneself into the machine. This is how most commentators think relying on intuitions as evidence works in this case (see Hewitt, 2010; Weijers, 2014; Rowland, 2017). Later in the article I am going to examine further examples in more detail to show that this understanding is virtually universal.

Bernard Molyneux has put forward a definition of the intuition dogma that is free of the two ambiguities I have just discussed:

[intuitions] are standardly *treated* as evidence of their contents, whether or not it is right to do so (Molyneux, 2014, p. 441).

The contents are then, of course, used as evidence for and against philosophical theories. Molyneux calls this view “descriptive evidentialism” (as he contrasts it with *normative* evidentialism, according to which intuitions *are* evidence). Admittedly, the term has not gained much popularity in the literature. Nevertheless, as I find Molyneux’s definition more accurate, I am going to prefer his term over the more widespread “Centrality”. I will attempt to show that descriptive evidentialism (henceforth “DE”) is the assumption behind most assertions about intuitions being relied on, accounted for, appealed to, deferred to, trusted, invoked, captured, matched, accommodated, systematised, explained, employed, used or treated as evidence in ethics.

4 The nature of intuitions

What exactly are intuitions, according to proponents of DE? Recently Nevin Climenhaga has offered the following definition:

I take intuitions to be mental states that we find ourselves in when considering particular propositions. I take it that when one has an intuition that P:

- (i) it seems to one that P;
- (ii) this seeming is not the conscious result of an inference;
- (iii) this seeming is not the conscious result of an apparent memory that P, a sensorial experience as of P, or someone else's testimony that P. (Climenhaga, 2018, pp. 69–70)

This is a fairly general account. Some theorists want to be more specific and more exclusive. While virtually everyone agrees that intuition is a propositional attitude—that is some sort of relation between an agent and a proposition—there is a deal of controversy over what kind of propositional attitude it is. Three main competing options are: a kind of belief, a kind of disposition to believe and a *sui generis* attitude. None is without difficulties. Opponents of the first argue that certain probability puzzles, like the Monty Hall problem, show it is not only possible to have an intuition that p without believing that p , but also to have an intuition that p while believing that not- p , which does not bode well for the belief theory. Opponents of the second option often argue it fails to account for the occurrent and episodic nature of intuitions (Pust, 2000, pp. 39–43). Opponents of the third deny that intuitions must always be occurrent—to think they are is to commit a “refrigerator-light fallacy”, that is to “confuse that which is always the case when you are looking with that which is always the case” (Earlenbaugh & Molyneux, 2009, p. 103).

Those who agree that intuitions are occurrent and episodic often add that they must also be spontaneous, or immediate—an intuitive episode cannot develop in a gradual way (Goldman & Pust, 1998, p. 179). Sometimes they also argue that they must be accompanied by a special phenomenology: there is *something it is like* to have an intuition, intuitions *seem true* is a particular way (Bealer, 1998, p. 207; Chudnoff, 2013, pp. 32–40). Bealer argues that any intuition used as evidence in philosophy has a specific kind of content: it “presents itself as necessary; it seems that things could not have been otherwise” (Bealer, 1999, p. 30). This is not true of any intuitive content. For example, in Newton's famous thought experiment we are asked to imagine a bucket partly filled with water, spinning in an otherwise empty space. It seems to us that water would creep up the side of the bucket, but not that this is *necessarily* the case. We are therefore dealing with what Bealer calls a “physical intuition”, which is not something philosophers typically rely on (Bealer, 1998, p. 205).

Intuitions are sometimes believed to be judgments generated by a special *faculty of intuition*, a sort of sixth sense. This view is often associated with the so-called ethical intuitionists, such as Henry Sidgwick, G. E. Moore or W. D. Ross. However there has been some controversies over how the faculty view should be interpreted and, consequently, whether different intuitionists actually subscribed to it (Crisp, 2002; Stratton-Lake, 2002). Another view associated with ethical intuitionism is that intuitive judgments are *self-evident*: they are justified simply by being understood, and no further justification for them can or needs to be offered. Unfortunately there seems to be little agreement on which particular judgments are self-evident, which means that neither this nor the faculty view can be fruitfully used as a criterion for distinguishing intuitions.

Some philosophers, like David Lewis and Peter van Inwagen, tend to be much more inclusive allowing inferential, and, for that matter, *any* beliefs to be classified as intuitions (Lewis, 1983, p. x; Van Inwagen, 1997, p. 309). In addition to restrictive accounts (like that of Bealer), moderately inclusive accounts (like that of Climenhaga) and broadly inclusive accounts (like that of Lewis) we can also distinguish idiosyncratic accounts that reject all the criteria listed above and introduce other criteria instead. One account of that sort is offered by John Rawls, who argues that while intuitions can be consciously inferred from other claims, they cannot be consciously inferred from *ethical principles* (Rawls, 1951, p. 183).

There is a lot more that can be said about the nature of intuitions according to different philosophers, however further discussion would be largely pointless. This is because in my view DE is false *irrespective of which account is adopted*, including the most liberal ones: it is not the case that the fact that p is non-inferential is used as evidence for p , it is not the case that the fact that p is partly non-inferential is used as evidence for p , it is not the case that the fact that p is believed is used as evidence for p , and so forth. Proponents of DE often argue that if intuitions are understood narrowly then perhaps they are not used as evidence in philosophy, however on a less restrictive understanding they clearly are used as evidence (Chalmers, 2014; Bengson, 2014; Stich & Tobia, 2016, p. 8). They accuse critics of DE like Cappelen of setting up a straw man: supposedly his way to question the practice of relying on intuitions in philosophy is to put a number of unreasonable qualifications on the nature of the intuitive. I think this is a misunderstanding: the reason why accusations like this are made is that DE is often conflated with something else—I will explain it in more detail later in the article.

5 The nature of evidence

So much for what exactly is used as evidence according to the view I will attempt to refute. We can now ask: what exactly does it mean to *treat something as evidence*, according to this view? In most cases proponents of DE do not specify how “evidence”, let alone “treating as evidence”, should be understood. Critics of DE tend to assume it is meant to be true irrespective of which specific theory of evidence is adopted and that it is possible to prove it false in a similarly theory-neutral way (Cappelen, 2012, pp. 11–12; Molyneux, 2014, p. 443). On the other hand, Climenhaga suggests we should be more specific. He proposes to understand DE along Bayesian lines:

E is evidence for T relative to background knowledge K iff $P(T|E\&K) > P(T|K)$
 – that is, E raises the probability of T relative to K. A person takes E to be evidence for T or uses it as evidence relative to K iff his conditional credence in T given E&K is greater than his conditional credence in T given K. (Climenhaga, 2018, p. 71)

This, however, strikes me as too author-oriented. First, philosophical writings are rarely framed as reports of their authors’ psychology, let alone estimates of fictions such as authors’ conditional credences in propositions. It is unclear how we can learn

much about what philosophers could believe given that something is true, based on what they write.

Second, on rare occasions when philosophers do comment on how their evidence influence their mental states, they do not necessarily confirm Climenhaga's view. Take William Lane Craig, who puts forward several arguments for the existence of God, most notably the so-called Kalam cosmological argument. Craig confesses that even if he were fully convinced that all his arguments were unsound, it would not diminish his belief in God one iota "because of the self-authenticating witness of God's Spirit who lives within him" (Craig, 2008, p. 46). Should we conclude that what Craig explicitly calls evidence for the existence of God is not taken by him to be evidence for the existence of God? Or perhaps that he must be wrong about his own beliefs?

Philosophers can be similarly attached to all sorts of philosophical beliefs. Think of Elizabeth Anscombe's remark about not wanting to argue with someone who thinks a judicial execution of an innocent person can be justified, as anyone who believes it "shows a corrupt mind" (Anscombe, 1958, p. 17). It seems perfectly possible to come up with reasons against executing an innocent person and treat them just as Craig treats his reasons to think God exists: as something that does not strengthen one's own belief that p , and yet supports p in one's published work.

There are also philosophers who do not appear to find their own arguments compelling in any way. William Lycan writes that if God offered him to bet on a doctrine he "would kill and die for" in his publications, he would not take the bet, even if the stake were only \$10 (Lycan, 2013, p. 115). Keith DeRose writes that if aliens who knew solutions to philosophical problems threatened him to destroy the Earth and entire humankind for not answering their philosophical question correctly, he would be more likely to go with the profession's majority view rather than the view he defends "when discussing the matter in a philosophical setting" (DeRose, 2017, pp. 267–269). Apparently this kind of scepticism is not the outcome of disbelieving one's premises or taking one's own arguments to be invalid. Lycan and DeRose might of course be wrong about what they would do in such outlandish circumstances. My point is not, however, that they must be right, but rather that there exists something they use as evidence when they do philosophy which is independent of how it influences their own beliefs.

In reply Climenhaga might argue that people like Craig, Lycan and DeRose are in fact outliers as most philosophers believe what they preach, and they believe it on the basis of their own evidence. This would mean that his definition can at least be used as a sort of rule of thumb for determining what is treated as evidence in philosophy. But even this is problematic. He mentions a distinction between private evidence and public evidence: the latter consists of reasons to accept a claim *offered in a public discussion* (Climenhaga, 2018, p. 98). For some reason, however, he is not troubled by the fact that his approach blurs the line between the two. I think this is a mistake. Philosophy is, after all, a public endeavour. A philosopher might believe that p for a number of reasons, and she might publish an argument which relies on p as one of the premises. This, however, does not mean that all her reasons to accept p are automatically used as philosophical evidence. To count as such, they must be *appealed to* in what is published. It might be the case that some philosophers believe certain things just because they find them intuitive. It might also be possible to find

some indication that they believe certain things just because they find them intuitive in their published work. But unless they offer the fact they (or someone else) find them intuitive as a reason to accept them, DE is not true.

One response to this problem might be to modify Climenhaga's definition by replacing the author with the reader: perhaps something is treated as evidence for a claim by a philosopher so long as it raises *the reader's* credence in the claim. This way private evidence could be kept out of the equation. The reader-centred approach also seems to make more sense of the fact that a philosophical argument is essentially a dialectical device: its primary point is to persuade whoever it is presented to, rather than to represent its author's internal thinking process.

But this proposal has serious flaws. First, just like philosophers' beliefs can remain intact by what they treat as evidence, their readers' beliefs can remain intact by what they are presented with as evidence. This might be due to irrationality, or for other reasons. For instance, I do not think that Zeno's paradoxes of motion make me any more likely to accept that motion does not exist. I believe that I can detect flaws that these paradoxes are based on, but, as I am not entirely sure whether I am right, they should have some influence on my view on the existence of motion. Moreover, when I first encountered the paradoxes I could not tell what was wrong with them, however I did not find Zeno's conclusion any more plausible. At least as far as I am concerned—but I suspect my case is not very odd—these arguments seem completely ineffectual. And yet clearly something is being used as evidence here.

Secondly, as Climenhaga points out, evidence on his view is *context-relative*: whether something counts as evidence always depends on one's background knowledge. This means that readers with different background knowledge cannot all rationally increase their credence in a claim by learning the same thing, and it is unclear which reader we should focus on. One might be tempted to overcome these difficulties by specifying we are only concerned with some sort of ideal reader with certain background knowledge, certain cognitive abilities, certain level of rationality etc. But this would not take us very far—if something is being treated as evidence for p when it raises the ideal reader's credence in p , then how can we know whether something raises the ideal reader's credence in p ? The answer must be either circular or unknowable.

Climenhaga's proposal and other possible Bayesian accounts can be characterised as instances of a *doxastic* view, according to which treating something as evidence is understood in terms of a relation between beliefs. The general idea behind this view can be expressed in the following way: p is treated as evidence for q if someone's belief that q is in some way based in their belief that p , where "based in" is understood broadly as causing, reinforcing, increasing the likelihood of etc. Note that most problems with Bayesian accounts that I have just described are also problems with doxastic accounts in general. This means we should probably abandon the doxastic picture of DE altogether: there seems to be no viable way of determining whether something is treated as evidence in philosophy in terms of how, if believed, it influences other beliefs.

To be fair to proponents of DE, not all of them are happy with the doxastic picture. Elijah Chudnoff suggests that it is more fruitful to understand treating as evidence in terms of a relation between an *experience* and a belief. He thinks that intuition is a

lot like perception—in fact it is “a form of intellectual perception” (Chudnoff, 2013, p. 1)—and to explain his idea it is useful to make an analogy with how we justify our perceptual beliefs. What evidence do we have for them? For example, what evidence do I have that there is a computer screen in front of me right now? The obvious answer is that I see the computer screen in front of me. But saying that seeing the screen justifies, or is evidence for, believing that there is a screen can be interpreted in several different ways. One of them would be doxastic: my belief that I see the screen justifies my belief that there is a screen in front of me. However epistemologists have identified a number of difficulties with this view (Lyons, 2016). Some of them argue that the best way to overcome these difficulties is to assume that it is my perceptual experience, *the seeing of the screen itself*, that directly justifies my perceptual belief. How exactly is this possible? Chudnoff’s answer is that it happens in virtue of the experience’s phenomenology: it is the way that it seems true that is connected to facts about what it represents. And, according to Chudnoff, what is true of perceptual experience is also true of intuition experience. This is why philosophers try to make their readers have a particular intuition experience, and this intuition experience is meant to serve as evidence for its own content—that is how “treating as evidence” in DE should be understood.

Does the experiential account of treating something as evidence fare any better than the doxastic account? It is doubtful. Later in the article I am going to explain how Chudnoff’s proposal leads to the idea that certain elements of philosophical writing are meant to trigger intuition experiences, and argue that this view is untenable.

But if the doxastic and experiential accounts are rejected, what are we left with? Instead of understanding treating something as evidence in terms of its psychological effects one can try to understand it in terms of a *logical* relation between propositions. After all, if there is one thing that all philosophers do to defend their views, it is *making arguments*. Why not simply take premises of an argument as something treated as evidence, and its conclusion as something it is meant to be evidence for? In other words, treating as evidence can be understood as synonymous with *inferring*. Inferences can, of course, be valid or invalid—we do not need to assume that philosophy is free of logical errors. They can also be of different types: deductive, inductive or abductive. But, in any case, they always link propositions, not mental states. Focusing on arguments themselves rather than on mental states of people who deal with the arguments seems to capture the phenomenon of philosophical evidence in a simpler and more straightforward way.

That said, it is important to stress that the inferential account is not irreconcilable with either doxasticism or experientialism. Perhaps what constitutes evidence *on a basic level* is a belief, or an experience. Perhaps it is something mind-independent, like a state of affairs. Whatever it is, it can be translated into an inference, and this inference is eventually expressed in natural language. The doxastic and experiential accounts of evidence in DE should therefore be dismissed only to the extent they refer to evidence that is not translatable into an inference that can be identified in a philosophical text. This restraint is dictated simply by the public nature of philosophy.

One can now ask: how do we go about testing DE, thus understood, as a hypothesis about philosophical practice? The most obvious solution would be to pay attention to linguistic means used to express the inferences in a text. That is we should look for

expressions like “so”, “therefore”, “hence”, “thus”, “it follows that”, “if—then”, “for”, “as”, “because”, “indicates that”, “suggests that”, “makes it plausible that”, “due to”, “is the reason why”, “for that reason”, “is a reason to think that”, “by virtue of”, “as a result of”, “accounts for”, “explains”, “on the basis of”, “thanks to”, and synonymous. If DE is true, on the one side of such connective we should be able to find the fact about a proposition’s intuitiveness, which would be expressed by phrases like “intuitively”, “it is intuitive that”, “there is an intuition that”, “it seems that”, “it appears that”, “it strikes me that”, “it is non-inferentially believed that” and so forth. On the other side of the inference-indicator we should be able to find the proposition itself. For example, something like “it seems that p , therefore p ”, or “the fact that p is intuitive suggests that p ” in a text would clearly support DE.

Here it might be objected that it is not impossible to rely on intuitions as evidence without mentioning it in the text. Perhaps the practice of relying on intuitions is so transparent and universally accepted that philosophers do not need to make it explicit. For example, Bealer writes that “is is truistic that intuitions are *used* as evidence (or reasons) in our standard justificatory practices” (Bealer, 1999, p. 30). If he is right, we probably should not expect philosophers to state the obvious. I think this objection needs to be taken seriously—later in the article I am going to explain how this “tacit” version of DE should be understood, and how it should be tested.

6 “The method of cases”

Proponents of DE typically argue that the practice of relying on intuitions is best exemplified by what they call “the method of cases”. Common instances include Searle’s Chinese Room, Putnam’s Twin Earth, Chalmers’s zombies, Nozick’s utility monster, Burge’s arthritis-in-the-thigh, Gettier cases, Frankfurt cases, trolley cases, Thomson’s violinist, Lehrer’s Mr Truetemp, Foot’s transplant surgeon, Jackson’s Mary the colour scientist or Kripke’s Gödel the thief.

But what is it that they all have in common? When explanation of any kind is given, virtually everyone agrees on three elements: there is the case itself, there is one particular judgment that the case is supposed to “elicit” or “trigger”, and there is a theory, or a generalisation, that the judgment is meant to be evidence for, or against. The description of the case is often, but not always, characterised as *a thought experiment*. The judgment is usually, but not always, characterised as *an intuition*. Here I set aside the intuition-free accounts of the method (such as Machery, 2017) and only focus on the more common, intuition-oriented ones (such as Malmgren, 2011 or Pust, 2019).

Let us take a closer look at one of the most prominent examples: the so-called trolley problem, first introduced by Philippa Foot (Foot, 1967), and later developed by Judith Jarvis Thomson in her two seminal articles (Thomson, 1976, 1985). I am going to focus on the two versions that for some reason have received most attention. The first is what I will call the bystander case. It is a modification of the original scenario described by Foot, in which a tram driver is about to hit and kill five people on the main track unless he turns the tram onto a sidetrack and kills one person. In Thomson’s new version it is not the driver, but a bystander that faces the dilemma:

you have been strolling by the trolley track, and you can see the situation at a glance: The driver saw the five on the track ahead, he stamped on the brakes, the brakes failed, so he fainted. What to do? Well, here is the switch, which you can throw, thereby turning the trolley yourself. Of course you will kill one if you do. (Thomson, 1985, p. 1387)

The other is what I will call the footbridge case:

you are standing on a footbridge over the trolley track. You can see a trolley hurtling down the track, out of control. You turn around to see where the trolley is headed, and there are five workmen on the track where it exits from under the footbridge. What to do? Being an expert on trolleys, you know of one certain way to stop an out-of-control trolley: Drop a really heavy weight in its path. But where to find one? It just so happens that standing next to you on the footbridge is a fat man, a really fat man. He is leaning over the railing, watching the trolley; all you have to do is to give him a little shove, and over the railing he will go, onto the track in the path of the trolley. (Thomson, 1985, p. 1409)

The first scenario is supposed to elicit the judgment that it is morally permissible to throw the switch and the second one that it is morally impermissible to push the fat man off the footbridge. How about generalisations that these judgments are supposed to undermine, or support? Some have suggested that the difference between the two somehow corresponds to the difference between utilitarian and rights-based ethics. For example, Joshua Greene argues that the apparent clash is “Kant versus Mill, all in one neat little puzzle” (Greene, 2013, p. 116). However this has little to do with the points Thomson is trying to make. In her article she simply takes it for granted that utilitarianism is a flawed moral theory: people have moral rights, and “rights trump utilities” (p. 1404). The bystander judgment might be in line with utilitarianism and the footbridge judgment might not, however Thomson is only interested in explaining the difference in terms of how different utility-trumping rights are violated, not violated or waived in both cases.

Immediately after having introduced the bystander scenario she makes it clear that the bystander judgment is meant to serve as a counterexample to “Killing one is worse than letting five die”, that is the principle defended by Foot in her discussion of the original version of the problem. If the bystander can throw the switch, then we have a situation when killing one is not worse. The footbridge scenario is in turn meant to provide a counterexample to “it is not morally required of us that we let a burden descend out of the blue onto five when we can make it instead descend onto one”. If the footbridge judgment is true, then we have a situation when this is exactly what is morally required of us.

Proponents of DE usually agree it primarily applies to “contemporary philosophy”, or “analytic philosophy”, however many point out that the method of cases has been in use since antiquity. The favourite example seems to be Plato’s discussion of justice in Book 1 of *The Republic*. This case is somewhat more problematic to interpret for the same reason any of Plato’s dialogues is problematic to interpret: the relation between the views presented by different characters and the author’s views is not always obvious. Moreover, the characters in *The Republic* are not discussing justice as such, but

rather the poet Simonides's beliefs about justice. I will assume that the standard interpretation, according to which Socrates's criticism of Simonides's definition of justice, "truth-telling and paying back what one has received from anyone", expresses Plato's view. Here is the famous counterexample:

I mean, for example, as everyone I presume would admit, if one took over weapons from a friend who was in his right mind and then the lender should go mad and demand them back, that we ought not to return them in that case and that he who did so return them would not be acting justly—nor yet would he who chose to speak nothing but the truth (Plato/Shorey, 1937, p. 11, 331c)

Any account of justice needs to take this fact into consideration, which means that there must be something wrong with the definition. In both cases, we can identify the description, the judgment and the generalisation undermined by the judgment. Proponents of DE argue that what makes the judgment special is the fact that it is intuitive. According to them, if the trolley judgments or Socrates's judgment were not intuitive, the whole exercise would be pointless: there would be no evidence against Foot's thesis about killing and no evidence against Simonides's theory of justice, respectively. We take these judgments to be true because they just *seem true*, and then we reject the generalisations as inconsistent with the judgments.

7 The justification interpretation

What can be wrong with the methodological picture I have just outlined? One objection raised by critics of DE like Max Deutsch is that "philosophers argue for their judgments about thought experiments and cases" (Deutsch, 2015, p. xvi). This means that we are expected to accept these judgments on the basis of *arguments*, not on the basis of the judgments' intuitiveness. I largely agree with Deutsch on this point, however I think his choice of words might be somewhat misleading. After all, the objective of *arguing for* a claim is typically to convince someone that the claim is true, however here we are dealing with claims that are probably already taken to be true by the interlocutors. Cappelen points out that judgments about cases often constitute "assumptions that in a typical non-philosophical context would be accepted by the conversation partners without a demand for further justification" (Cappelen, 2012, p. 189), and since philosophy is about questioning everything, philosophers often try to find justification that is not demanded in a typical non-philosophical context. So instead of saying that philosophers *argue for* judgments about cases, I think it would be more accurate to say that they *provide justification* for them, or that they *explain what makes them true*, or they *back them up with evidence* (which has nothing to do with their intuitiveness).

For example, Thomson backs up her judgments that it is permissible to throw the switch and that it is not permissible to push the fat man off the bridge with the following principle:

it is not morally required of us that we let a burden descend out of the blue onto five when we can make it instead descend onto one if we can make it descend

onto the one by means which do not themselves constitute infringements of [stringent] rights of the one (Thomson, 1985, p. 1409)

Similarly, Plato backs up the judgment that returning the weapons would not be just by stating that “friends owe it to friends to do them some good and no evil” (Plato/Shorey, 1937, pp. 13, 332a)—this is Cephalus’s reply to Socrates, referring to Simonides’s beliefs, which reveals that either Simonides is being inconsistent, or returning the weapons cannot be classified as “paying back what one has received”.

If the justification interpretation is correct, it seems that nothing turns on whether judgments about cases are intuitive: all that matters is whether their justification is sound. Proponents of DE often overlook the justification of judgments about cases, or sometimes even explicitly deny it is present in the text (for example, see Gutting, 1998, p. vii). This seems to lead them astray: they mistakenly conclude it is the intuitiveness of the judgments that is intended to support them.

8 The abductive interpretation

There is, however, a DE-friendly way of accounting for what Deutsch calls arguments for judgments about cases. Proponents of DE can argue that judgments like “it would be wrong to return the weapons to one’s friend who has gone insane” are not inferred from principles like “friends owe it to friends to do them some good and no evil”—it is the other way around. We are dealing with a sort of *inference to the best explanation*, or *abduction*, from the former to the latter (I am going to use the two terms interchangeably, which is not universally accepted—see Mackonis, 2013). On this account judgments about cases serve as independently attested data to be explained by theories. For example, when someone argues that human activity is the best explanation for the crop circles, they take it for granted that crop circles exist—apparently because crop circles have been observed. Similarly, when Plato argues that “friends owe it to friends to do them some good and no evil” is the best explanation of why it would be wrong to return the weapons, he takes it for granted that it would be wrong to return the weapons—apparently because it has been *intuited*. In both cases we have an independent source of knowledge of the facts we are attempting to explain: observation and intuition, respectively.

But this response runs into serious difficulties. First, offering the best explanation for data typically involves acknowledging, more or less explicitly, that there exist other explanations and demonstrating they are inferior, according to certain criteria. For example, Peter Lipton writes that “better explanations explain more types of phenomena, explain them with greater precision, provide more information about underlying mechanisms, unify apparently disparate phenomena, or simplify our overall picture of the world.” (Lipton, 2001, p. 106) To stick with the crop circles example: proposing human activity as the best explanation for their existence typically involves acknowledging that extraterrestrial intervention has been proposed as an alternative explanation. This, however, does not resemble what philosophers do while discussing the paradigm cases. For example, Plato does not mention, or even hint at, any alternative explanations of why it is wrong to return the weapons to a friend and does not argue

that they are worse in any of the respects mentioned by Lipton. This suggests that Plato is not engaging in abductive reasoning.

Secondly, it is worth examining the wording of the relevant passages. As Deutsch admits, one needs to be careful here, as expressions like “explains”, “accounts for”, “is the reason for” or “because” can be used to represent both a deductive and abductive inference (Deutsch, 2015, pp. 96–97). I agree with Deutsch that the inference-language can be ambiguous, however some cases are still fairly clear-cut. For example, there seems to be little room for interpretation of how Plato uses the word “for” / γάρ (“for he believes that friends owe it to friends...”/ τοῖς γὰρ φίλοις οἶεται ὀφείλειν τοὺς φίλους...). This is not how one would normally present this claim were it meant to serve as the best explanation of why we should not return the weapons. I am not suggesting that a word like “for” or “because” would be completely out of place in a presentation of an abductive inference, however it should not appear as its main indicator. Consider: “Crop circles exist because they were created by humans”. This sentence might look suitable in a concluding section of a discussion of what best explains crop circles, but as a standalone statement it simply would not work. For the same reason Plato’s sentence does not work as an abduction-indicator.

Third, philosophers sometimes come across conflicted judgments about cases, and the way they deal with the conflict shows they do not understand them as independently attested data to be explained by theories. For example, throughout her paper Thomson reports presenting her friends with different versions of the trolley scenario and asking them about their opinions. In most cases there is a consensus. Sometimes, however, her judgment differs. There is a version of the bystander scenario in which in order to throw the switch it is necessary to cross a patch of land that belongs to the person on the sidetrack, or to use his nail file, in both cases without the owner’s permission. To Thomson’s interlocutors diverting the tram in this situation seems permissible, but Thomson herself “does not find it obvious”. In another scenario, the person on the sidetrack, which has been unused for years, is a convalescent at a local hospital, having a picnic lunch. He was invited there by a city mayor, who had promised him no trams would ever be diverted onto the sidetrack. Unexpectedly, a tram is about to hit and kill five people on the main track, unless someone turns it towards the convalescent, and the only person who can do so happens to be the mayor himself. To Thomson’s “great surprise”, her interlocutors thought it would be permissible for the mayor to throw the switch in this situation as well.

Here are some possible ways of dealing with the judgment discrepancy, assuming the DE-friendly abductive interpretation is correct: one could conclude it is impossible to proceed as there is no clear intuition-data to explain; one could try to come up with different generalisations for different sets of judgments, one could try to find out which judgment is more widespread or more strongly intuitive; one could try to argue that someone’s faculty of intuition—if there is such a thing—was impaired or malfunctioning in some way. As it turns out, Thomson does none of these things. Instead, she looks into *reasons* to accept and reject judgments about cases. In the patch of land/nail file case, she argues that her interlocutors must be correct as “the rights which the bystander would have to infringe here are minor, trivial, non-stringent-property rights of no great importance” (p. 1411). In the city mayor case, she is not likely to change her mind straight away. She believes her interlocutors assume that

breaking one's promise does not infringe a stringent right, or at least a right not stringent enough to override the exemption allowing to sacrifice one in order to save the five. Thomson remains unconvinced: it seems clear that in order to resolve the disagreement it would be necessary to examine reasons *behind reasons* to accept the judgment, that is reasons to think that breaking one's promise is too trivial to override the exemption. In both cases, Thomson believes judgements about moral permissibility of particular actions should be accepted or rejected on the basis of an *argument*, not on the basis of whether they are intuitive to anyone. The intuitiveness of judgments seems completely irrelevant.

Fourth, philosophers sometimes make comments about what on the abductive interpretation serves as the best explanation that the advocates of this interpretation must find baffling. For example, Thomson says she does not “find it clear why there should be an exemption for, and only for, making a burden which is descending onto five descend, instead, onto one” (p. 1408). Note that on the abductive interpretation this comment does not make much sense. Thomson should find it perfectly clear why there should be an exemption: the exemption thesis accounts for a number of judgments about different versions of the scenario. However she does not appear to think that the judgments are something that can justify, or support the exemption thesis. Rather, it is the other way around. We are then left with the exemption thesis that is far from obvious or self-evident, which means it needs to be supported by some further facts. Thomson says one such fact is that we are dealing with “something that is *already* a threat to more, and thus something that will do harm *whatever* [the bystander] does” (ibid.), but this can only serve as a *partial* justification. She feels she does not have enough evidence to justify the exemption thesis—hence her perplexity.

Fifth, philosophers sometimes *change* their judgments about cases over time and the way they do so does not bode well for the abductive interpretation. For example, in 2000s Thomson had come to the conclusion that after all it was *not* permissible to throw the switch and divert the tram in the bystander case. How was it possible? On the DE-friendly abductive view, Thomson's intuition about the case must have changed, or she must have decided something had been wrong with her ability to intuit the correct answer, or perhaps she must have learnt that people's intuitions about the scenario were different than she had previously thought—in any case, there must have been some sort of turnaround, failure or misunderstanding concerning someone's intuitions, which serve as an independent source of data to be explained by a theory. The problem is this is nowhere near how Thomson actually explains her change of mind. She says she was persuaded by Alexander Friedman, who argued that since she had first presented the problem, nobody—herself included—had been able to offer a satisfactory account of what makes it permissible to throw the switch. According to Friedman this is because there is no such account to discover. On the other hand, we have a good reason to believe that it would be wrong to divert the tram onto the sidetrack: “it is intuitively plausible that negative duties really are weightier than positive duties.” (Thomson, 2008, p. 363).

Here the objection might be that the word “intuitively” indicates that we are dealing with a situation in which it is impossible to account for all intuitions and one intuition (“negative duties are weightier”) simply trumps another (“it is fine to throw the switch”), but overall it is still true that intuitions are treated as starting premises in

abductive arguments. But this response is problematic for a number of reasons. First, why did it take Thomson several decades to realise that “negative duties are weightier” is intuitive and therefore has to be treated as some sort of explanandum? How did it not occur to her in 1976 or 1985? This seems highly implausible. Another possibility would be that Thomson did not find the proposition intuitive in the past, but this conjecture seems even more far-out: surely, if it were the case, she would have at least flagged it up in her article. Secondly, if Thomson or Friedman are trying to somehow weigh two intuitions against each other, why are they not invoking any criteria for solving this kind of conflict? Why exactly is one intuition supposed to override the other? Is it because it is more intuitive, or for some other reason?

There is much more to be said about the idea of sacrificing intuitions for the sake of preserving other intuitions, and I will return to this problem later in the article. As for using the word “intuition” and its cognates in one’s first order philosophical practice: it is undeniable that many philosophers do use this kind of terminology, however it never signals anything resembling DE. Cappelen offers several DE-unfriendly accounts of philosophical intuition-talk and argues that they always explain it much better than DE (Cappelen, 2012, pp. 25–93). His list might not be exhaustive, but his overall conclusion seems right.

9 The noninferential interpretation

Another DE-friendly interpretation of considerations that I call evidence for judgments about cases has recently been proposed by Chudnoff. According to it, these considerations are neither inferred from the judgments, nor the judgments are inferred from them: there is simply no inference-relation between the two. Rather, the considerations *enable* the judgments. Here is how Chudnoff understands the difference:

If you infer c from $p_1 \dots p_n$, then your justification for believing c is constituted by your justification for believing $p_1 \dots p_n$. Say your justification for believing in the principle of mathematical induction is constituted by the testimony of a textbook. Then in the inference case your justification for believing the formula is partly constituted by testimony. If consideration of $p_1 \dots p_n$ enables your intuition that c , then your justification for believing c need not be constituted by your justification for believing $p_1 \dots p_n$. Rather, it is constituted by your intuition and whatever background information it draws on. Say your justification for believing that $4 + 2$ is 6 is constituted by the testimony of a textbook. You learned this in school and just haven’t thought about it since. Nonetheless, in the intuition case your justification for the formula need not be partly constituted by testimony. It is important not to assume that considerations used to enable an intuition are thereby incorporated into the background information drawn on in the intuition. (Chudnoff, 2021, p. 147)

According to Chudnoff, an analogy can be drawn between enabling intuition and enabling perception. Consider the phenomenon of multistable perception: certain images can depict different things, depending on which way they are looked at. For example, in a popular image known as “My wife and my mother-in-law” one can see a

young woman facing away or a left profile of an old woman—but not both at the same time. We can imagine, argues Chudnoff, someone who can only see the old woman in the picture, and someone else telling him that the old woman’s nose is the young woman’s jawline, the old woman’s mouth is the young woman’s necklace etc. This consideration would make the first person see the young woman, but it would not constitute evidence that there is a young woman in the picture. The evidence would be *the very experience of seeing the young woman*, together with whatever background information it makes use of. Similarly, the role of considerations like “friends owe it to friends to do them some good and no evil” could be to merely make a proposition like “it would be wrong to return the weapons” seem true, without justifying it.

However there are strong reasons to think that Chudnoff’s interpretation is not correct. First and foremost, a consideration like “friends owe it to friends...” simply does not seem to enable the experience of finding “it would be wrong to return the weapons” intuitive, and the same can be said about other case judgments and their respective considerations. Note that proponents of DE typically ignore considerations like “friends owe it to friends...” when they discuss judgments about cases (see Pust, 2019 or Stich & Tobia, 2016). On Chudnoff’s account, this should lead to some sort of fatal miscommunication between proponents of DE and their readers: certain judgments are constantly pronounced to be epistemically special in virtue of being intuitive, but the reader cannot find them intuitive, as there is nothing in the text to enable their intuitiveness. But no such miscommunication happens: nobody seems to accuse philosophers like Pust of arbitrarily calling certain judgments “intuitions” without offering any justification.

Secondly, if the non-inferential interpretation of the relation between considerations like “friends owe it to friends...” and judgments like “it would be wrong to return the weapons” is correct, why do philosophers routinely use “inferential” language to describe it? I have pointed out that words like “for”, “because”, “as” etc. in the original texts refer to the said relation. Note how unnatural it would be to say “there is a young woman in the picture *because* the old woman’s nose is the young woman’s jawline, etc.”. If Chudnoff’s analogy between perception and intuition is valid, it should also be unnatural for Plato to say “it would be wrong to return the weapons *because* friends owe it to friends...”—and yet it is precisely how Plato formulates his sentence.

Having said that, I believe Chudnoff is on to something when he argues that philosophers do rely on intuitions to make things “more vivid” —he is only wrong to think that this practice has something to do with DE. Later in the article I am going to explain how I think intuitions are used as tools of discovery and tools of clarification—something that often is, but should not be conflated with what is typically meant by “relying on intuitions”.

10 Common ground

I have argued that judgments about cases are typically *backed up with evidence*. Suppose I am wrong and there is nothing to support the judgments in the text. Would that make DE plausible? I do not think it would. It would only mean that judgments about cases are unsupported and other claims are inferred from them. This tells us nothing

about whether the intuitiveness of unsupported claims plays any kind of justificatory role.

It is hardly surprising that philosophical arguments, or *any* arguments, for that matter, rest on unsupported premises. It simply follows from the fact that arguments cannot be infinitely long. Trying to support one's unsupported premises means one is only going to end up with another set of unsupported premises. However the fact that all arguments rest on unsupported premises does not mean that all arguments rest on intuitions, in the DE-sense. For example, I can start an argument with an unsupported claim that the distance between Tehran and Isfahan is shorter than the distance between Tehran and Shiraz. Does it mean I am using the fact that this claim is in some sense intuitive to support its content? Of course not. Most likely I am simply assuming this is something my readers already know, so I do not need to waste their time explaining why it is the case. Or, should they not know it, that the claim's truth and evidence in favour of it is quite uncontroversial and easy to look up, I can therefore expect the readers to take my word for it. Simply put, I am placing "the distance between Tehran and Isfahan is shorter" *in the common ground*.

Let us now ask: why not think about various philosophical starting premises in the same dialectical way—namely as something that does not need to be argued for in a particular text? Why not think it is the quality of being already accepted by the readers that makes various judgments suitable to start philosophical arguments with? Someone might reply that if we identify "intuition" with "something assumed to be already accepted by the readers", it would follow that intuitions are used as evidence. However, even putting aside the eccentricity of this usage, DE would still be false: on this account the fact that someone assumes p to be widely accepted is clearly not meant to be evidence for p .

The common ground interpretation has at least one clear advantage over DE: it explains the suspicious lack of explicit claims in the form of " p is intuitive, therefore p " or similar in the texts in question. As Deutsch points out, if DE were true we would expect perhaps not all, but at least some philosophers to conform to this pattern—but in fact none of them do (Deutsch, 2015, p. 97). Of course Deutsch's argument does not undermine the "tacit agreement" version of DE, and the difference between it and the common ground interpretation might be somewhat elusive. David Chalmers, who defends a form of DE, suggests to understand it in the following way:

Propositions in the common ground typically have a broadly inferential dialectical justification: it is just that this justification is in the background, stemming from how the proposition entered the common ground in the first place. Often the justification will be a testimonial or perceptual justification, deriving from previous communications or from external sources. As before, these dialectical justifications need not be explicitly articulated by the parties to a conversation; they merely need to be mutually recognized. By contrast, with intuitions as I am characterizing them, there need be no broadly inferential justification that the parties recognize; there will only be a broadly noninferential justification, perhaps associated with the obviousness of the claim in question. (Chalmers, 2014, p. 538)

I agree with Chalmers that mutual recognition of justification typically characterises propositions in the common ground, however his account seems too restrictive. Suppose a philosopher puts forward an argument whose starting premise is p . She assumes that all her readers accept p , but she is not sure *why* they accept it. This situation hardly vindicates DE, however it is reasonable to say that p is in the common ground. Or suppose that a philosopher puts forward an argument whose starting premise is p , but she thinks different readers are going to accept p for different reasons. Here, again, even though there is no one particular justification recognised by all parties, the common ground interpretation seems correct while DE clearly does not. Finally, suppose that a philosopher puts forward an argument whose starting premise is p and assumes all her readers accept p , but she cannot think of any reasons for p . This situation still does not confirm DE—in fact, it is not even consistent with DE, which states that a philosopher offers a particular reasons for p , namely that p is intuitive. In short, “common ground” is better understood as “assumed to be accepted by all parties for any reason, or even without an identifiable reason”, rather than “assumed to be accepted by all parties for the same reason (other than being intuitive)”. The tacit version of DE would in turn imply that something is assumed to be accepted by all parties on the basis of being intuitive—on this point Chalmers seems to agree.

It must be stressed that placing p in the common ground does not equal believing that p . One often starts with what one’s opponent’s already believe, without necessarily believing it oneself—familiar phrases like “for the sake of argument”, “I’ll grant you that”, “let’s assume that” etc. are often used in this context. The proponent of the argument may even disbelieve her own starting premises and be open about it—although keeping one’s attitude towards p to oneself is also perfectly consistent with putting p in the common ground. Simply put, what is in a philosopher’s common ground should be treated as independent of what she believes and whether she reveals what her beliefs are.

How can one decide between the common ground hypothesis and the tacit DE hypothesis in a given case? First, we can check whether p , which serves as a starting premise in an argument, is challenged *in a different text*—by a different or perhaps even the same author. If DE is true, we would expect the text to mention the consensus that the intuitiveness of p counts as evidence for p . Such text might not, of course, always be available. However we can also *imagine* p being challenged, and ask how likely the participants of the debate would be to bring up the alleged consensus. We can also ask: what considerations against p could possibly be offered and how reasonable would it be to weigh these considerations against the fact that p is intuitive? This test is, of course, far from perfect due to its hypothetical nature. But, as I will try to show later in the article, it is reliable enough to assess DE.

As I mentioned, I do not believe that the common ground interpretation is correct with regards to the judgments I have discussed—I think both Thomson and Plato provide justification for their judgments. Perhaps I am wrong about this, or perhaps the same is not true about other paradigm judgments listed above. In any case, proponents of DE always face a double challenge: first, they need to show that judgments about cases constitute argumentative starting points, and, secondly, they need to show how the fact that judgments about cases are starting points supports DE. I do not think this challenge can ever be met, regardless of which alleged example of relying on

intuitions one takes up. Moreover, what I have just said about the so-called method of cases can be applied to any assertion about using intuitions as evidence in philosophy. Whenever DE commits one to identifying a proposition as an intuition that is being treated as evidence, two questions can be asked: is this proposition backed up with any evidence (other than the fact that the proposition is intuitive) in the text and if not, is it part of the common ground? DE implies that the answer to both questions is no. In my view, the answer to one them is always yes.

11 “*Prima facie*”

Proponents of DE might complain that I have just presented them with a false dilemma: it is possible to accept the justification interpretation and still maintain that intuitions are used as evidence of their contents. Perhaps intuitions are taken to be some sort of *defeasible* evidence, that is something that can then get confirmed or undermined—maybe even overridden—by further evidence. For example, it can be argued that Plato treats the intuitiveness of “one should not return the weapons” as defeasible evidence that one should not return the weapons, and this defeasible evidence is then further confirmed by the fact that “friends owe it to friends to do them some good and no evil”. Or perhaps Thomson treats the intuitiveness of “it is not permissible to use the nail file to throw the switch” as defeasible evidence that it is not permissible to use the nail file to throw the switch, and then this evidence is overridden by the fact that using the nail file without the owner’s permission does not violate a stringent right of a person. Philosophers like to talk about “*prima facie* reasons”, “*prima facie* objections”, “*prima facie* problems”, “*prima facie* doubts” or “*prima facie* counterexamples” —perhaps what they mean is this kind of defeasible evidence?

I do not think this account holds water. Several reasons to reject the abductive interpretation are also reasons to reject the *prima facie* interpretation. First, there is the lack of explicit inferences from “*p* is intuitive” to “*p*”. For example, Plato does not say that one should not return the weapons because friends owe it to friends to do them some good and no evil *and* because it is intuitive that one should not return the weapons. He says that one should not return the weapons because friends owe it to friends to do them so good and no evil, full stop. If Plato is appealing to the intuitive, why does he stay silent about this? And why does virtually everyone else stay silent?

It might be objected that even though philosophers are not explicitly appealing to two different sources of evidence for a particular claim, they can still hint at them by using particular words and expressions. Ethan Landes defends this position in his recent paper. He focuses on Edmund Gettier’s famous refutation of the justified true belief theory of knowledge, which, along with Thomson’s trolley cases or Plato’s discussion of justice, is often exhibited as a paradigm example of the method of cases in use. Gettier words his justification in the following way: “But it is equally clear that [Gettier’s judgment], for [Gettier’s reason to accept the judgment].” (Gettier, 1963, p. 123). Landes acknowledges that the reason is given in the text—the word “for” leaves little room for doubt. But he also thinks that by using the phrase “equally clear” Gettier *additionally* justifies his judgment by pointing to the fact it is intuitive (Landes, 2020, p. 10). I do not think, however, that this reading of Gettier’s sentence is even

remotely plausible. Consider the statement: “It is clear he does not have a PhD in philosophy, for he has never heard about Kant or Hegel.” How reasonable would it be to interpret it as “He does not have a PhD in philosophy for it is intuitive that he does not have a PhD in philosophy *and also* because he has never heard about Kant or Hegel”? Surely “clear” in my sentence indicates that I take my reason for believing that someone does not have a PhD in philosophy to be strong, and consequently I think my conclusion is well-established. It does not refer to the fact that I take the alleged intuitiveness of my conclusion to support its content. The same goes for Gettier’s sentence, and the point can be generalised to all other judgments about cases. The wording of relevant passages simply does not favour the *prima facie* interpretation.

12 Discovery vs justification

At this point one might ask: if this is not what philosophers mean by “*prima facie*”, what do they mean? If they do not think that the intuitiveness of *prima facie* claims matters in terms of justification, why are they even talking about them? I doubt whether there is one uniform way of using the expression “*prima facie*” in philosophy, however I think there is a plausible interpretation of how it is often used, which is both incompatible with DE and allows a role for intuitions to play in philosophical inquiry.

An analogy may be helpful here. Imagine a detective to whom a strong intuition occurs: it seems to her that one of the suspects has committed the crime, but she has no idea why. She decides to follow the intuition and pursue a certain line on inquiry, in the course of which she is able to collect evidence that reveals the suspect to be the culprit: fingerprints, DNA samples, CCTV recordings, witness testimony etc. The evidence is then presented at the trial. The detective does not treat her own intuition as worthless: she thinks it indicates that there is evidence to be found somewhere, and not much evidence to be found elsewhere. She does, however, treat it as worthless *in court*: arguing that someone had an intuition that someone else was guilty cannot help convict anyone, and she is perfectly aware of this fact. In this metaphor the detective’s intuition is analogous to our intuitions about philosophical cases; fingerprints, DNA samples etc. are analogous to whatever philosophers justify those judgments with, and the trial is analogous to a typical philosophical debate. Philosophers take what seems true to us and try to find out what, if anything, backs it up. Whatever they think backs up is treated by them as evidence. Whatever they back up with evidence is the *prima facie* claim.

One might argue that if the detective thinks her intuition indicates there is (court-compliant) evidence to be found somewhere and not much elsewhere, she is clearly using her intuition *as evidence for its content*: the fact that something seems true to her indicates that it is in fact true, or likely to be true. I do not disagree with this point, however it needs to be stressed that we are talking about *non-public evidence* here. Just like it would be bizarre to appeal to this kind of non-public evidence in court, it would be bizarre to appeal to this kind of non-public evidence in a philosophical debate. Intuitions are simply not used to publicly support or undermine philosophical views, even though they can, in a sense, lead one to discover what is used to support and undermine philosophical views. Philosophers often assume that when it

seems to us that something counts as justice, knowledge, reference etc., we are on to something—that is, there probably are good reasons to think it actually counts as justice, knowledge, reference etc. They then try to discover what those reasons are, and publish their findings.

Moreover, my analogy can easily be modified to eliminate treating intuitions even as non-public evidence while still engaging in essentially the same practice. Suppose that the detective does not really trust her intuition, but, for whatever reason, she still decides to follow it, which leads her to discover evidence. The same can be true of a philosopher: she can give priority to intuitive judgments in the process of examining reasons behind them without ever treating the judgments' intuitiveness as evidence of its content. The reason for prioritising these judgments can be simply that they are more *interesting* than random judgments which nobody finds plausible. There is a significant overlap between what is intuitive and what is believed, and we are naturally more curious about what we believe: we want to know why we believe it and whether we are justified in believing it.

It is sometimes argued that DE is supported by the fact that respectable philosophical theories of justice, knowledge, reference etc. generally accommodate our intuitions about what counts as justice, knowledge, reference etc. To anyone who rejects DE, the objection goes, this must look like a surprising coincidence (Climenhaga, 2018, pp. 79–80). On my view, however, there is no coincidence. Philosophers often pay more attention to our intuitions, however this does not mean they treat intuitions as evidence—at least not as *public evidence*. This explains why, for example, Thomson reports asking her friends about different trolley scenarios. Perhaps she assumes her friends must be on to something when they make their verdicts, perhaps she merely finds the verdicts more attractive to explore. In any case, she is not *justifying* claims with the fact that her friends make them, or find them intuitive, which is what DE implies. Or consider how Thomson describes her change of mind about the bystander case in her 2008 article: she writes that many philosophers for many years have focused on judgments like “it is permissible to throw the switch” and “it is impermissible to push the fat man off the bridge” and strove to find good justification for them, but failed (Thomson, 2008, p. 363). Because of this failure we should turn to nonintuitive judgments, such as “it is impermissible to throw the switch” and see if they can be justified. Whether something counts as a good justification has nothing to do with the fact that it is intuitive, however it is still true that philosophers often prioritise intuitive judgments in their investigations.

A similar point has been made by R. M. Hare. He argues that moral philosophers often appeal to what he calls “the opinions of the ordinary man”. Plato’s “one should not return the weapons” is one of his examples. According to one interpretation, these opinions are used as data to be explained by moral theories. But if this is true, it follows that philosophers are “merely being conservative or conventionally-minded or just stupid” (Hare, 1972, p. 124–5). In any case, they are engaging in a terrible kind of reasoning, as intuitions are clearly a very bad guide to moral truth. Fortunately, writes Hare, there is another, more plausible way of understanding the practice:

in spite of the fact that the opinions of the ordinary man have in themselves no probative force in moral philosophy, a due respect for them may lead us to

understand its problems better. They do not supply an argument, but they make us look for one. (p. 134)

Since Hare wrote it in 1970, much seems to have changed for worse. The argument from received opinion used to be one of several interpretative possibilities, today it has become the prevalent view. This might look somewhat surprising if we compare contemporary philosophy of philosophy with contemporary philosophy of science, which uses a well-established distinction between the context of discovery and the context of justification. The former has to do with, roughly, actual thinking processes behind the creation of new scientific ideas and theories, and the latter with what is used to evaluate those ideas and theories in the scientific community. In an anecdote often invoked to illustrate the distinction Friedrich August Kekulé is led to discover the ring structure of benzene by dreaming about a snake seizing its own tail (Kekulé, 1890/1958, p. 22). Even though Kekulé's dream played an important role in the discovery, it would be strange to suggest that dreams can be treated as evidence in chemistry. Perhaps Kekulé thought that the content of his dreams carried some evidential weight, perhaps he did not: as far as scientific justification is concerned, this is beside the point. The hypothesis about the ring structure of benzene had to be tested by standards that were independent of contents of anyone's dreams.

A number of ways of understanding the discovery vs justification distinction have been proposed since Hans Reichenbach introduced it in 1938: it can refer to two processes distinct in time, to the *process* of discovery and *methods* of justification, to something that can be analysed empirically and something that can be analysed logically, or to something still different (see Hoyningen-Huene, 2006). The version that I propose to apply in metaphilosophy centres on theory validation: according to it, evidence in the context of philosophical justification is simply something suitable to support or undermine a theory *in a philosophical debate*. There might also exist evidence in the context of philosophical discovery: something that can be relied on in a creative process, but not something that can be used to justify a theory in a philosophical community. And DE, as I understand it, always refers to treating intuitions as evidence *in the context of justification only*.

Distinguishing between relying on something as evidence in the context of discovery and relying on something merely as a working hypothesis—without treating it as evidence in any sense—is an intricate psychological matter. For example, I am not sure whether the way that Kekulé relied on his dream resembled the way that the detective relies on her intuition in my first thought experiment, or perhaps the way she relies on intuition in the second one. I have similar doubts with regard to the way many philosophers seem to rely on intuition. As I do not need to solve this issue to make my case against DE, for simplicity's sake I am going to refer to both kinds of practice as “using as evidence in the context of discovery”.

13 Clarification and persuasion

Just like Thomson's trolley judgments, Plato's “weapons” judgment is probably not intuitive *by accident*. According to his argument, fulfilling certain obligations, such

as one's obligation to do good to one's friends, can require not giving back what one owes—which conflicts with the definition of justice put forward by Simonides. To clarify, Plato introduces his thought experiment. As it *seems obvious* that returning the weapons would be wrong, we can easily understand how the premises support the conclusion. We can imagine a situation in which it is not obvious that someone should not give back what she owes, and yet it still follows from the fact that she should do good to her friend. Had Plato decided to use such judgment, his argument would have become more difficult to comprehend, but its substance would not have changed. We can say that Plato is relying on an intuition *as a clarification device*, which is very different from using an intuition as evidence.

A related, but distinct function of appeal to intuitions has to do with *persuasion*. In addition to making himself clear, Plato probably tries to make his readers believe that justice is not what Simonides's definition suggests, and the intuitiveness of "one should not return the weapons" helps him achieve this goal. It is well known that impeccable arguments are neither necessary nor sufficient for successful persuasion. One can accept the premises of an argument as undeniably true and the logic of it as perfectly valid and yet still refuse to accept the conclusion. For this reason, it may be wise to avoid conclusions that seem false, and opt for ones that seem true. This, of course, is not always possible, but when it is possible philosophers often take advantage of the opportunity, which gives us yet another type of philosophical practice that can be confused with DE.

Using intuitions as clarification devices and using them as persuasion devices often go hand in hand, but there is no necessary connection between the two. Consider the Monty Hall problem mentioned earlier. A game show host offers you a choice between three gates. Behind one of them there is a prize, two others are empty. You pick your gate, then the host opens one of the empty gates and asks if you would like to change your original choice. To most of us it seems false that the probability of winning increases after switching. We think it is just obvious that it does not matter whether we switch or not, the probability is $1/2$ either way. There are many ways of explaining of why this is not the case. For example, one can make use of formal probability calculus. Another solution would be to slightly modify the original scenario. Suppose that instead of three gates there are a hundred of them. Everything else stays the same: there is only one prize behind one gate, and the host knows which one it is. You pick your gate, the host then opens ninety-eight empty ones and asks if you would like to change your original choice. In this case most people immediately understand that switching increases their chances of winning: clearly the probability is $1/100$ if you stick, and $99/100$ if you switch. After all, if you are lucky with your first guess ($1/100$ chance that you are), the alternative gate would be a randomly selected empty one. And if you are unlucky ($99/100$ chance that you are), the alternative gate would be the winning one. But if this is so, then in the three gate version of the game the probability of winning must rise from $1/3$ to $2/3$ after switching.

Jason Rosenhouse, who has spent years teaching probability theory using the Monty Hall example, writes that "students who are totally unpersuaded by elaborate probability calculations or arguments based on Bayes' theorem typically cry uncle at this point" (Rosenhouse, 2009, p. 39). Elaborate calculations do not necessarily fail to clarify, however they do fail to persuade. For this reason, a mathematician whose aim

is to actually change people's minds about the chances of winning after switching may *rely on an intuition* by preferring the hundred gate scenario over other ways of explaining the puzzle. But it would be absurd to think that the (correct) intuition that it is advantageous to switch in the hundred gate version is used as evidence that it is indeed advantageous to switch, or that the (mistaken) intuition that it is not advantageous to switch in the three gate version is used as evidence that it is not advantageous to switch. These intuitions tell us nothing about whether switching is a good idea, and they are treated accordingly by mathematicians. In my view, philosophers are not any different in this respect: they often use intuitions to persuade without using them as evidence.

14 Four ambiguities

So far I have argued that the claim that philosophers rely on intuitions is ambiguous in at least four ways. First, there is the evidence versus clarification/persuasion device ambiguity. Then within the evidence interpretation there is the propositional content versus mental state ambiguity. Then within the mental state interpretation there is the evidence for its content versus evidence not for its content ambiguity. And finally, within the evidence for its content interpretation there is the context of discovery vs context of justification ambiguity. The last reading is what I call DE and what I argue against in this article.

It is undeniable that in some sense philosophers do rely on intuitions, however what is typically meant by “philosophers rely on intuitions” is DE, and DE is false. It might be objected that my approach is too stringent: perhaps the commitment to DE is not as widespread as I suggest it is. Let us explore this possibility. First, why should we reject the clarification/persuasion device reading? The main reason seems to be that “philosophers rely on intuitions” is often used interchangeably with “philosophers rely on intuitions *as evidence*”, and it would make little sense to talk about *evidence* in this context: helping someone understand or trying to persuade someone that something is the case is different from giving evidence for why something is the case. Other common synonymous expressions are “philosophers account for intuitions with their theories” and “philosophers construct their theories by appealing to intuitions”—and they seem equally incompatible with the clarification/persuasion device interpretation, according to which intuitions are clearly not any kind of building blocks or raw material of theories. Moreover, it is often claimed that intuitions are indispensable in philosophical theorising, that it is impossible for philosophers not to rely on intuitions in one way or another, etc. However under the clarification/persuasion device interpretation there is nothing indispensable about intuitions. It may be helpful to appeal to them while presenting a theory, but nothing beyond that: giving up on such appeals does not change the substance of argumentation.

The problem with the propositional content reading seems fairly straightforward: if “intuitions” were to mean exclusively “propositional contents of intuitions”, then what would be the point of singling out this kind of propositional content? I agree that propositions that merely *happen* to be intuitive are often used as evidence, but I also think that propositions that happen not to be intuitive are often used as evidence, in

very much the same fashion: both can serve as starting or intermediate premises of philosophical arguments. Recall the example I used at the beginning of this article: “propositions formulated by carbon-based life forms are used as evidence in philosophy”. This statement is not only literally true, it may be true about all philosophical evidence. And yet it sounds odd—this is because being formulated by carbon-based life forms is a property that stems from a historical contingency, not from anything *methodologically salient*. After all, in principle non-carbon-based life forms or sophisticated machines seem perfectly capable of formulating the exact same propositions.

Perhaps it might be objected that if “evidence” is limited only to starting premises, and if a very broad account of “intuitions” is adopted, then all propositions used as evidence in philosophy would count as intuitive, their intuitiveness would be methodologically salient and yet it would not be treated as evidence for those propositions. As I mentioned, some philosophers, like van Inwagen and Lewis, argue that intuitions can be identified simply with beliefs or opinions. On this account, it would be true that philosophers generally try to start their arguments with intuitions—that is with what they think is already accepted by their readers. Surely there is little point in offering an argument which starts with premises that the addressees of the argument are going to reject straight away. Is it, however, really what those who argue that philosophers rely on intuitions have in mind? This is highly implausible. First, when they specify what they mean by “intuition”, they practically always opt for a narrower account—typically one that at least involves non-inferentiality. On the narrower account many starting premises of philosophical arguments are not intuitions. Secondly, those who argue that philosophers rely on intuitions typically also argue that there is something uniquely philosophical about this practice, and clearly there is nothing uniquely philosophical about trying to start an argument with premises already accepted by its intended recipients. It is a common feature of arguments as such, not just philosophical arguments, and it would not be reasonable to suppose that this fact is widely overlooked.

Let us now turn to the “evidence not for its content” interpretation. The main reason to disqualify it is the typical choice of examples that illustrate the thesis: we are constantly reminded that the practice of relying on intuitions is best exemplified by trolley cases, Plato’s discussion of justice etc. This clearly suggests that intuitions are meant to be used as evidence for their contents. If the fact that “one should not return the weapons” is intuitive can help refute Simonides’s theory of justice, it is only because it is used as evidence that one should not return the weapons, and the fact that one should not return the weapons is incompatible with the claim that justice is “truth-telling and paying back what one has received from anyone”. The same goes for all other examples. Moreover, a number of philosophers make this point explicit. For example, Christopher Daly writes that “those who appeal to intuitions take an intuition that p to provide *prima facie* evidence that p ” (Daly, 2015, p. 11). Similar claims have been made by Norbert Paulo (2020, p. 334), Brian Weatherson (2003, pp. 19–20), Alvin Goldman and Joel Pust (1998, p. 181), or James Andow (2017, p. 184).

Finally, there is the evidence in the context of discovery reading. One problem with it is that there is only so much we can learn about this kind of evidence from studying philosophical material. When Kekulé first published his findings about the structure of benzene, he did not mention dreaming about a snake seizing its own tail, as it was—and still is—considered inappropriate to include detailed information concerning one’s

own creative process in scientific publications. Philosophy is not much different in this respect: even if it is more acceptable to include such information, the information is often not there. This means that what is eventually published can be the outcome of many different ways of thinking. It is not impossible to learn something about how a given argument came about, but this often requires reaching beyond strictly philosophical publications to sources such as interviews, letters, diaries, memoirs, private conversations etc. However those who argue that philosophers rely on intuitions hardly ever refer to such sources. The entire evidence that Plato relies on an intuition is to be found in *The Republic*, the entire evidence that Thomson relies on intuitions is to be found in her papers on the trolley problem, and so on.

It might be objected that I am now hoist by my own petard as I myself argue that philosophers use intuitions as evidence in the context of discovery without appealing to extraphilosophical sources I have just mentioned. There is, however, an important difference between what I do and what I argue is hard to explain on the context of discovery interpretation of “philosophers rely on intuitions”. I think some *limited* evidence for my claim can be found in philosophical publications. For example, I have pointed out that Thomson discusses reasons to think that it is impermissible to push the fat man off the bridge, but never mentions any reasons to think it is permissible to do so. I think this counts as evidence that in her thinking process she has not given each option a fair hearing, but rather focused only on reasons behind the intuitive one. I have also argued that the way philosophers use the expression “*prima facie*” in their first-order philosophical practice—as opposed to their metaphilosophical claims—does not indicate engaging in anything resembling DE, while it may well indicate engaging in relying on intuitions in the context of discovery. Overall I think that *solely on the basis of what can be found in philosophical sources* we can conclude that the context of discovery hypothesis is always a better explanation of what philosophers do than DE. This, however, is far from offering a full-blown defence of the context of discovery hypothesis, which would require a careful examination of extraphilosophical material. In contrast, those who argue that philosophers rely on intuitions never seem to think that appealing to such material would be suitable, which suggests they are talking about using evidence in the context of justification.

Moreover, the way the intuition thesis is typically worded leads to the same conclusion. We hear that intuitions in philosophy are like observations in science, that theories are judged to be acceptable if they capture intuitions, that refuting a theory amounts to showing it has counterintuitive implications, etc. None of these expressions make much sense on the context of discovery reading. Perhaps it could be objected that I am interpreting at least some of the expressions uncharitably—for example, “rejecting a theory because of its counterintuitive implications” might simply be a shorter and less precise way of saying “rejecting a theory because of reasons discovered by examining its counterintuitive implications”. This, however, seems too much of a stretch. Analytic philosophers pride themselves on being exceptionally meticulous. Sometimes their devotion to rigour is even seen as a flaw: apparently it makes academic texts lengthy, dry, tedious and generally unreadable. How plausible is it that in one particular case philosophers universally decide to prioritise conciseness over precision? It is true that claims about relying on intuitions are sometimes little more than passing comments, but even then making it clear that one is not referring to DE

does not require a lot of effort. Given the confusion that results from using less precise language, there is simply too much to lose and too little to gain.

Having said that, I concede that on rare occasions claims like “philosophers rely on intuitions” and similar do not necessarily reveal the commitment to DE. I have already cited Hare whose discussion of the problem makes this point clear. Another example, perhaps a slightly less straightforward one, is a point made by Thomas Nagel:

Given a knockdown argument for an intuitively unacceptable conclusion, one should assume there is probably something wrong with the argument that one cannot detect – though it is also possible that the source of the intuition has been misidentified. If arguments or systematic theoretical considerations lead to results that seem intuitively not to make sense, or if a neat solution to a problem does not remove the conviction that the problem is still there, or if a demonstration that some question is unreal leaves us still wanting to ask it, then something is wrong with the argument and more work needs to be done. Often the problem has to be reformulated because an adequate answer to the original formulation fails to make the sense of the problem disappear. (Nagel, 2002, pp. x–xi)

I believe this passage not only can, but most likely should be interpreted in a DE-unfriendly way. According to Nagel it is not the fact that one has an intuition that speaks directly against the argument, but rather the fact that one has an intuition indicates there must be something wrong with the argument, even though one is currently unable to *detect* what it is. All this is compatible with the denial of DE: taking on the argument in a philosophical publication would require identifying and describing the mistake, merely asserting that something is counterintuitive and therefore must be false would be dismissed as a flawed kind of reasoning, not worthy of consideration. Moreover, if the same mistake is not identified by means of intuition, it changes nothing as far as justification goes. One might object that Nagel is putting too much faith in intuition: perhaps many arguments for counterintuitive views are perfectly sound and what he recommends is a wild-goose chase. But whether it is or not is an epistemological question that has no bearing on DE. When he says that philosophers should trust their intuition, and that many of them—himself included—do trust it, he does not endorse DE in any way.

Nagel seems to understand “relying on intuitions” as “relying on intuition-states as evidence of their contents in the context of discovery”. Occasionally one can come across other DE-unfriendly readings. For example, Timothy Williamson dismisses the idea of not relying on intuitions as a “non-starter” by pointing out that all reasoning, philosophical and non-philosophical alike, must begin with unsupported premises (Williamson, 2018, p. 63). Perhaps he simply identifies “intuitions” with “starting premises”. Perhaps what he has in mind is slightly different: “intuitions” are “widely shared beliefs”, whose contents are suitable to serve as starting premises. In any case, his reading of “philosophers rely on intuitions” is a truism nobody objects to. Nevertheless, interpretations like that of Williamson or Nagel seem exceptional. As I have tried to show, in most cases we can find a more or less explicit commitment to DE.

15 Reflective equilibrium

“The method of cases” might not be the full story of how intuitions are used as evidence in philosophy, assuming DE is true. Some argue that while the method is only concerned with intuitions about particular situations, philosophers also appeal to intuitions about more general or abstract principles (Sosa, 2009, p. 10; Strevens, 2019, p. 3). Pust gives an example of taking the fact that “a suitably formulated consequentialist theory” is “intuitive in itself” to be evidence for consequentialism (Pust, 2000, p. 11). However if we allow such general intuitions to play an evidential role, it soon turns out they can clash with intuitions about particular cases, and that trade-offs are necessary. Proponents of DE typically think there is a widespread method of confronting intuitive claims of different levels of generality with each other in order to achieve coherence: the so-called method of reflective equilibrium. Here is how Norman Daniels outlines the idea:

The method of reflective equilibrium consists in working back and forth among our considered judgments (some say our “intuitions,” though Rawls (1971), the namer of the method, avoided the term “intuitions” in this context) about particular instances or cases, the principles or rules that we believe govern them, and the theoretical considerations that we believe bear on accepting these considered judgments, principles, or rules, revising any of these elements wherever necessary in order to achieve an acceptable coherence among them. (Daniels, 2020)

This can seem perfectly compatible with DE: first certain judgments are taken to be supported, explicitly or implicitly, by the fact they are intuitive, and then philosophers work back and forth among them until some sort of equilibrium state is reached. Perhaps what I have described as offering reasons for and against intuitive judgments is part of the method, and should not be held against DE.

How do we determine whether philosophers actually engage in reflective equilibrium seeking? The question is not easy to answer. “Reflective equilibrium” may sound like a well-defined philosophical term of art, but in fact different philosophers interpret it differently, and in many cases it is hard to tell how it is understood. For example, is the method coherentist or foundationalist? The dominant view is that it is the former: no proposition is immune from revision in the process of seeking coherence. However Pust argues that on any reasonable interpretation we are dealing with a method “within which the process of justification is linear and stops with intuitions” (Pust, 2000, p. 13). Secondly, for a judgment to be treated as an input, is being intuitive a necessary or a sufficient condition, or perhaps neither? Rawls makes it clear that intuitiveness is merely one of a number of properties required: considered judgments also need to remain uninfluenced by certain emotional states, self-interest or threats, they need to be made by people with a degree of intelligence, certain kind of understanding of how human interests can conflict etc. (Rawls, 1951, pp. 178–183; Rawls, 1971, p. 47) However all these additional requirements are often neglected in contemporary discussions of the method. The status of principles, rules and theoretical considerations mentioned by Daniels is also far from clear: are they all suitable to be used as starting points because they also are, in some sense, intuitive? Third, what do we mean by “intuitive”? As I

mentioned, Rawls's understanding of the term is substantially different from that of most contemporary theorists. Fourth, if Rawls is right that additional requirements are needed, what exactly are they? Should we accept Rawls's list, modify it, or perhaps come up with an entirely different one? Stefan Sencerz outlines four broad answers to this question (Sencerz, 1983, pp. 83–90)—and it is far from clear which, if any, is correct. Fifth, does “seeking equilibrium” refer to a creative process that ultimately leads to the publication of a philosophical work, or perhaps to the work's content? Sixth, is there a separate stage of inquiry when a set of starting points is determined and another one when the process of confronting them with each other takes place? Or perhaps the two are intertwined? Seventh, is the process carried out by an individual, or is it dialogical? Daniels claims that both uses are possible, however the common understanding seems to be the individualist one. Eighth, are we talking about the so-called narrow or wide reflective equilibrium? According to Rawls, the latter does not involve investigating “principles people would acknowledge and accept the consequences of when they have had an opportunity to consider other plausible conceptions and to assess their supporting grounds” (Rawls, 1974, p. 8). Some theorists clarify which of the two they are talking about, but many others ignore the distinction. Ninth, are we talking about a method of ethics, or a method of philosophy in general? Rawls thinks it is the former, and this is how the method is typically understood today. However something closely resembling Rawls's idea can be found for example in Nelson Goodman's work on the justification of deductive and inductive inferences (Goodman, 1955). Finally, to answer all the above, what are the canonical texts that we should we turn to? For example, should we treat Rawls's later contributions (1971, 1974) as more authoritative than his early paper (1951) just because the latter does not use the term “reflective equilibrium”? And do all three describe essentially the same practice, or, as it is suggested by Pust, three different ones? Again, no straightforward answer can be given.

Whether something counts as an example of using the method depends on how one answers the questions listed above. I do not deny that for some sets of answers the result would be positive: in a sense, philosophers sometimes seek reflective equilibrium. I do, however, deny that such practice ever vindicates DE. Let me illustrate this point. David Benatar argues that coming into existence is always a serious harm, which means that procreation is always immoral and that humanity should die out as quickly as possible. As he is aware, most people find his conclusions deeply counterintuitive. He is not, however, averse to the idea of accounting for intuitions. Rather, he argues that he has to dismiss some intuitions so that other intuitions can be preserved. The gist of his argument is that there is a certain asymmetry of pain and pleasure: the presence of the former is bad and the presence of the latter is good, but while the absence of the former is good, the absence of the latter is *not* bad. The claim that coming into existence is always a harm is the best explanation of this asymmetry. It can of course be asked why we should accept the asymmetry in the first place, but Benatar thinks no one who cares about preserving intuitions should go down that road:

Of course, there are various ways of rejecting asymmetry, but the least implausible way would be by denying that absent pleasures are ‘not bad’ and claiming instead that they are ‘bad’. This would commit us to saying that we *do* have a

(strong?) moral reason and thus a presumptive duty, based on the interests of future possible happy people, to create those people. It would also commit us to saying that we can create a child for that child's sake and that we should regret, for the sake of those happy people whom we could have created but did not create, that we did not create them. Finally, it would commit us not only to regretting that parts of the earth and all the rest of the universe are uninhabited, but also to regretting this out of concern for those who could otherwise have come into existence in these places.

Matters become still worse if we attempt to abandon asymmetry in another way—by claiming that absent pains (...) are merely 'not bad'. That would commit us to saying that we have no moral reason, grounded in the interests of a possible future suffering person, to avoid creating that person. We could no longer regret, based on the interests of a suffering child, that we created that child. Nor could we regret, for the sake of miserable people suffering in some part of the world, that they were ever created. (Benatar, 2006, p. 204)

It may seem that Benatar is trying to account for a number of intuitions and DE is a perfectly accurate description of his practice. Why would I deny it? Recall that DE comes in two varieties: the explicit and the tacit. On the explicit account, Benatar should be saying things like "it is intuitive that we have no moral duty to create happy people based on those people's future interests, therefore we have no such duty". However he never frames his inferences this way, which leaves us with the tacit variety. According to it, Benatar takes the claim that we have no such duty to be supported by its own intuitiveness, and he thinks his readers treat it in a similar way—therefore he does not need to make it clear in the text.

But this is highly implausible. First, Benatar stresses that he treats claims like "there is no duty to create happy people" as his starting points *as they are widely accepted*, and that the fact they are widely accepted *is not a reason to think they are true* (p. 36). This shows that he explicitly rejects DE as a description of his own practice if "intuition" in DE is to be understood as "widely accepted". How about other understandings, like "psychologically non-inferential", or "spontaneous", or "seeming true in a phenomenologically distinct way"? If Benatar thought that the judgment's non-inferentiality etc. is a reason to accept it, we would expect him to state that in the very same passage. The fact he does not strongly favours the common ground interpretation.

Moreover, Benatar is aware that none of his starting points is shared *universally*. He mentions that certain utilitarians may be inclined to argue that there is, after all, a kind of duty to procreate. No names are named in the book, but it is not hard to find them. For example, Torbjörn Tännsjö writes that "to the extent that we add creatures living lives worth living, our ambition to replenish the universe not only is part of our quest for meaning, but also means that we comply with our duties as moral agents" (Tännsjö, 2002, p. 355). If DE were true, we would expect Tännsjö to argue that the intuitiveness of "there is no duty to create happy people" is either no evidence that there is no duty to create happy people—despite what is commonly believed—or that it is trumped by some evidence to the contrary. But Tännsjö says neither of these things. Sure enough, he tries to explain why his conclusion seems false to many of us, but he

spends no time explaining why the fact it seems false is no reason to dismiss it, or a relatively weak reason to dismiss it. This is most likely because hardly anyone takes it to be such reason—there is no need to undermine the consensus as the consensus does not exist.

It might be objected that Tännsjö and his fellow utilitarians do not in fact disagree with Benatar's starting premise. This is because the premise refers to creating happy people *for those people's sake*, while the utilitarian claim refers to creating people *for the sake of the sum of all happiness*, whose maximisation is the sole goal of morality. However even if this objection is valid, and even if all philosophers on Earth agree with Benatar's starting premises, there are still no grounds to think that these premises are meant to be tacitly supported by the fact they are intuitive. It is perfectly legitimate to question and defend them by appealing to all sorts of considerations. For example, one might argue that fulfilling the alleged duty would turn many women into constantly pregnant procreation-machines, which would violate their autonomy; or that we cannot have duties towards non-existent beings whose existence is dependent on the fulfilment of the alleged duty. On the other hand, it would be strange to argue that the duty to create happy people does not exist because we tend to form this judgment spontaneously, or because making it is accompanied by a distinct phenomenology, or because we cannot think of why it is true, etc. There is something deeply unphilosophical about assertions of this kind, and they simply would not count as reasons in a serious conversation. However DE implies not only that such reasons can be respectable, but that they constitute the most obvious and universally recognised evidence.

16 Conclusion

I have argued that the idea of relying on intuitions in ethics, as it is commonly understood, is deeply mistaken, and that there is a glaring gap between what ethicists do and what ethicists think they do. This conclusion has some serious implications, not the least about the relevance of empirical studies on intuitive judgments often carried out under the label of “experimental philosophy”, but I am not going to discuss them here. Regardless of how significant this error is for various disciplines and research programmes, simply as a metaphilosophical misunderstanding it is embarrassing enough to be promptly abandoned. On the other hand, the view that ethicists rely on intuitions has not come from nowhere. It is false, but not delusional: there are true views that can be, and often are, equivocated with it. If I am correct, both sides of the debate have some work to do: those who subscribe to the intuition-orthodoxy should fundamentally revise their views, and the dissenters should pay more attention to the role that intuitions actually play in philosophical methodology.

Declarations

Conflict of interest The author has no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long

as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Andow, J. (2017). A partial defence of descriptive evidentialism about intuitions: A reply to Molyneux. *Metaphilosophy*, 48(1–2), 183–195.
- Anscombe, E. (1958). Modern moral philosophy. *Philosophy*, 33(124), 1–19.
- Audi, R. (1993). Ethical reflectionism. *The Monist*, 76(3/1), 295–315.
- Bealer, G. (1998). Intuition and the autonomy of philosophy. In M. R. DePaul & W. Ramsey (Eds.), *Rethinking intuition. The psychology of intuition and its role in philosophical inquiry* (pp. 201–239). Rowman & Littlefield.
- Bealer, G. (1999). A theory of the a priori. *Philosophical Perspectives*, 13, 29–55.
- Bedke, M. S. (2008). Ethical intuitions: What they are, what they are not, and how they justify. *American Philosophical Quarterly*, 45(3), 253–269.
- Benatar, D. (2006). *Better never to have been. The harm of coming into existence*. Oxford University Press.
- Bengson, J. (2014). How philosophers use intuition and 'intuition.' *Philosophical Studies*, 171, 555–576.
- Boyd, R. (1988). How to be a moral realist. In G. Sayre-McCord (Ed.), *Essays on moral realism* (pp. 181–228). Cornell University Press.
- Brandt, R. (1979). *A theory of the good and the right*. Oxford University Press.
- Cappelen, H. (2012). *Philosophy without intuitions*. Oxford University Press.
- Chalmers, D. (2014). Intuitions in philosophy: A minimal defense. *Philosophical Studies*, 171, 535–544.
- Chudnoff, E. (2013). *Intuition*. Oxford University Press.
- Chudnoff, E. (2021). *Forming impressions. Expertise in perception and intuition*. Oxford University Press.
- Climenhaga, N. (2018). Intuitions are used as evidence in philosophy. *Mind*, 127(505), 69–104.
- Craig, W. L. (2008). *Reasonable faith. Christian truth and apologetics*. Crossway.
- Crisp, R. (2002). Sidgwick and the boundaries of intuitionism. In P. Stratton-Lake (Ed.), *Ethical intuitionism: Re-evaluations* (pp. 56–75). Oxford University Press.
- Daly, C. (2015). Introduction and historical overview. In C. Daly (Ed.), *The Palgrave handbook of philosophical methods* (pp. 1–30). Palgrave Macmillan.
- Daniels, N. (2020). Reflective equilibrium. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/sum2020/entries/reflective-equilibrium/>.
- DeRose, K. (2017). *The appearance of ignorance. Knowledge, skepticism, and context* (Vol. 2). Oxford University Press.
- Deutsch, M. (2015). *The myth of the intuitive. Experimental philosophy and philosophical method*. MIT Press.
- Earlenbaugh, J., & Molyneux, B. (2009). Intuitions are inclinations to believe. *Philosophical Studies*, 145, 89–109.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5–15.
- Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, 23(6), 121–123.
- Goldman, A., & Pust, J. (1998). Philosophical theory and intuitional evidence. In M. R. DePaul & W. Ramsey (Eds.), *Rethinking intuition. The psychology of intuition and its role in philosophical inquiry* (pp. 179–197). Rowman & Littlefield.
- Goodman, N. (1955). *Fact, fiction & forecast*. Harvard University Press.
- Greene, J. (2013). *Moral tribes. Emotion, reason and the gap between us and them*. Penguin.
- Griffin, J. (1988). *Well-being: Its meaning, measurement, and moral importance*. Clarendon.
- Gutting, G. (1998). "Rethinking intuition": a historical and metaphilosophical perspective. In M. R. DePaul & W. Ramsey (Eds.), *Rethinking intuition. The psychology of intuition and its role in philosophical inquiry* (pp. 3–13). Rowman & Littlefield.

- Hare, R. M. (1972). The argument from received opinion. In R. M. Hare (Ed.), *Essays on philosophical method* (pp. 117–135). University of California Press.
- Hoyningen-Huene, P. (2006). Context of discovery versus context of justification and Thomas Kuhn. In J. Schickore & F. Steinle (Eds.), *Revisiting discovery and justification. Historical and philosophical perspective on the context distinction* (pp. 119–131). Springer.
- Hewitt, S. (2010). What do our intuitions about the experience machine really tell us about hedonism? *Philosophical Studies*, 151(3), 331–349.
- Kagan, S. (2001). Thinking about cases. *Social Philosophy and Policy*, 18(2), 44–63.
- Kamm, F. (2007). *Intricate ethics. Rights, responsibilities, and permissible harm*. Oxford University Press.
- Kauppinen, A. (2015). Moral intuition in philosophy and psychology. In J. Clausen & N. Levy (Eds.), *Handbook of neuroethics* (pp. 169–183). Dordrecht: Springer.
- Kekulé, A./ Benfey, O. T. (1890/1958). August Kekulé and the birth of the structural theory of organic chemistry in 1858. *Journal of Chemical Education*, 35 (1), 21–23.
- Landes, E. (2020). The threat of the intuition-shaped hole. *Inquiry*, 1–26.
- Lewis, D. (1983). *Philosophical papers* (Vol. 1). Oxford University Press.
- Lipton, P. (2001). Is explanation a guide to inference? A reply to Wesley C. Salmon. In G. Hon & S. S. Rakover (Eds.), *Explanation. Theoretical approaches and applications* (pp. 93–120). Springer.
- Lycan, W. (2013). On two main themes in Gutting's *What Philosophers Know*. *The Southern Journal of Philosophy*, 51(1), 112–120.
- Lyons, J. C. (2016). Experiential evidence? *Philosophical Studies*, 173, 1053–1079.
- Machery, E. (2017). *Philosophy within its proper bounds*. Oxford University Press.
- Mackonis, A. (2013). Inference to the best explanation, coherence and other explanatory virtues. *Synthese*, 190(6), 975–995.
- Malmgren, A. (2011). Rationalism and the content of intuitive judgements. *Mind*, 120(478), 263–327.
- McMahan, J. (2013). Moral intuition. In H. LaFollette & I. Persson (Eds.), *The Blackwell guide to moral theory* (pp. 103–120). Wiley.
- Molyneux, B. (2014). New arguments that philosophers don't treat intuitions as evidence. *Metaphilosophy*, 45(3), 441–461.
- Nado, J. (2017). Demythologizing intuition. *Inquiry*, 60(4), 386–402.
- Nagel, T. (2002). *Mortal questions*. Cambridge University Press.
- Nozick, R. (1974). *Anarchy, state, and utopia*. Basic Books.
- Nozick, R. (1989). *The examined life. Philosophical meditations*. Simon & Schuster.
- Paulo, N. (2020). The unreliable intuitions objection against reflective equilibrium. *The Journal of Ethics*, 24, 333–353.
- Plato/Shorey, P. (1937). *Plato's Republic, Books I-V*. Harvard University Press.
- Pust, J. (2000). *Intuitions as evidence*. Routledge.
- Pust, J. (2019). Intuition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/sum2019/entries/intuition/>.
- Rawls, J. (1951). Outline of a decision procedure for ethics. *The Philosophical Review*, 60(2), 177–197.
- Rawls, J. (1971/1999). *A theory of justice*. Harvard University Press.
- Rawls, J. (1974). The independence of moral theory. *Proceedings and Addresses of the American Philosophical Association*, 48, 5–22.
- Rosenhouse, J. (2009). *The Monty Hall problem. The remarkable story of math's most contentious brain teaser*. Oxford University Press.
- Rowland, R. (2017). Our intuitions about the experience machine. *Journal of Ethics and Social Philosophy*, 12(1), 110–117.
- Shaw, W. H. (1980). Intuition and moral philosophy. *American Philosophical Quarterly*, 17(2), 127–134.
- Sencerz, S. (1983). Moral intuitions and justification in ethics. *Philosophical Studies*, 50(1), 77–95.
- Sinnott-Armstrong, W., Young, L., & Cushman, F. (2010). Moral intuitions. In J. Doris (Ed.), *The moral psychology handbook* (pp. 46–73). Oxford University Press.
- Sosa, E. (2009). A defense of the use of intuitions in philosophy. In D. Murphy & M. Bishop (Eds.), *Stich and his critics* (pp. 101–112). Wiley.
- Stich, S., & Tobia, K. (2016). Experimental philosophy and the philosophical tradition. In J. Sytsma (Ed.), *A companion to experimental philosophy* (pp. 5–21). Wiley.
- Stratton-Lake, P. (2002). Introduction. In W. D. Ross (Eds), *The right and the good* (pp. ix-1). Oxford University Press.

- Strevens, M. (2019). *Thinking off your feet. How empirical psychology vindicates armchair philosophy*. Harvard University Press.
- Tännsjö, T. (2002). Why we ought to accept the Repugnant Conclusion. *Utilitas*, 14(3), 339–359.
- Thagard, P. (2010). *The brain and the meaning of life*. Princeton University Press.
- Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, 59(2), 204–217.
- Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, 94(6), 1395–1415.
- Thomson, J. J. (2008). Turning the trolley. *Philosophy & Public Affairs*, 36(4), 359–374.
- van Inwagen, P. (1997). Materialism and the psychological-continuity account of personal identity. In J. Tomberlin (Ed.), *Philosophical perspectives 11. Mind, causation and world* (pp. 305–19). Wiley.
- Weatherson, B. (2003). What good are counterexamples? *Philosophical Studies*, 115(1), 1–31.
- Weijers, D. (2014). Nozick's experience machine is dead, long live the experience machine! *Philosophical Psychology*, 27(4), 513–535.
- Williamson, T. (2018). *Doing philosophy. From common curiosity to logical reasoning*. Oxford University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.