

How To Do Moral Philosophy

Paul Boswell
University of Michigan, Ann Arbor

Arguments

Every academic field employs a favored set of analytical tools in pursuit of truth, and philosophy is no exception. Anthropology uses case studies, literature close reading, macroeconomics simplified formal models, and a wide variety of fields (including philosophy, when it finds them useful) use classical statistical inference, real analysis, and higher mathematics.

What's unique among philosophers is our preoccupation with formal argumentation.¹ Even when critiquing someone else's argument for a particular conclusion, philosophers will put forward their critique *as an argument* itself. This is because philosophers and logicians have developed a fairly technical understanding of what arguments are and how they work, one which allows us to more easily check whether a conclusion has been definitively established as true, and, if not, where exactly the reasoning has gone wrong. So learning how to read, evaluate, and write arguments is the central skill of philosophy: philosophy is a game in which truth is the object and the rules are given by the logic of argumentation. If you can't do those three things passably well you won't get much out of a philosophy course, just as you're unlikely to win in football if you only have a hazy understanding of the rules.

¹Well, maybe not wholly unique. Mathematicians use arguments in pretty much *exactly* the same sense as philosophers do, except that they are called "derivations" or "proofs" and differ in a few important but technical details.

So what is an argument? Here's the technical definition:

Argument A set of propositions, exactly one of which is interpreted as a *conclusion*, and the rest of which (if any) are interpreted as *premises*.

In formal work the individual premises of an argument are listed with the conclusion appearing at the very bottom, usually with a line appearing between the last premise and the conclusion. The whole idea of an argument is that the premises are supposed to provide *support* for the conclusion, in that they all together provide *reason to believe* that the conclusion is true. But this is not part of the definition of an argument. Here is a *bona fide* argument:

1. Paul is wearing two shoes.
2. Paris is the capital of France.

3. So, a Republican will be elected US President in 2032.

Again, this *is* an argument. It's just a really bad one.

But I'm getting ahead of myself, since the definition of "argument" above contains another technical term I need to define:

Proposition A proposition is something a person could in principle *believe* or *desire* to be the case and which is either true or false (but not both).

Both of the premises and the conclusion in the bad argument above are propositions.² (Check both parts of the definition.)

²You might at first think that there's no truth *now* about outcome of the the Presidential election in 2032, but it is probably better to say that you just don't now *know* whether it's true or false. For, consider the sentence "Robert Kennedy dies in 1968." That was true in 1968 just as it is today, because that sentence describes the world as it actually came to pass. So a clairvoyant who said that sentence in 1868 would *also* be saying something true, and for the exact same reason. This should lead us to say that "A Republican will be elected President in 2032" is (now) either true or false, even though we don't know whether it's true and it's the sort of truth over which we have control, at least in a collective sense. If you still disagree with me here, you can just replace "is either true or false" with "was, is, or will be either true or false" in the above definition.

Exercise. How about these — are these propositions?

- “Shut the door!”
- “Chocolate is delicious.”
- “This sentence is false.”

So far this is all by way of clarifying the notion of an argument; we have not said anything about what makes an argument good or bad. Partially this is because it’s hard to say exactly what makes an argument a good one, and there are many ways for an argument to be bad. But in philosophy, most really good arguments have this technical property called “validity”:

Validity An argument is *valid* just in case it is impossible for all of its premises to be true and its conclusion false. Equivalently, an argument is *valid* just in case necessarily, if all of its premises are true, then its conclusion is true.

We can also say that the conclusion *follows* from the premises of a valid argument, or that they *entail* it. Intuitively, a valid argument is one whose premises *guarantee* the truth of its conclusion.³ That’s what’s so great about validity from an argumentative standpoint: if you give a valid argument for a conclusion, then your opponent *must* either accept the conclusion or disagree with one of the premises, on pain of inconsistency. They cannot say, “I agree with all your points, but I just don’t think they support your conclusion.”

Many good arguments are valid, but so are many bad arguments. An important class of bad arguments that are often valid are those that *beg the question* against an opponent of the conclusion.

There’s no generally accepted definition of “begging the question,” but it’s easy to point to an example. Suppose a theist is having a debate with an atheist and makes the following argument for the existence of God:

1. The Bible says that God exists.
2. The Bible is the inspired word of God.

³It’s important not to read this “guarantee” in causal terms, however. Verify that “The moon is made of blue cheese” entails “If grass is green, then grass is green.”

3. Nothing that is the inspired word of God can be false.

4. So, God exists.

This is a perfectly valid argument: if the Bible says that God exists and every sentence of the Bible is true, then God must exist. But it's not a good one. No one who disagreed with the conclusion of the argument at the outset would have any reason to accept the second premise, which directly entails the conclusion. After all, if the Bible is the inspired word of God, then He must be around to inspire it! So insofar as the goal of argumentation is to give someone genuine *reasons* to agree with you, you fail to achieve that goal if you smuggle the contentious conclusion directly into one of the premises. That's begging the question.

Note that for this reason, it's still a bad argument even *if* all the premises are true and the Bible is the inspired word of God. The theist must establish that conclusion on other grounds.

The Is-Ought Gap

So far this has just been a general introduction to argumentation. But this is a class in moral philosophy, and in this field a crucial complication arises out of the nature of the conclusions argued for — in particular, out of the fact that they include normative words like “should”, “ought”, “good”, and “unjust”. David Hume puts the problem this way:

In every system of morality, which I have hitherto met with, I have always remark'd, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surpriz'd to find, that instead of the usual copulations of propositions, *is*, and *is not*, I meet with no proposition that is not connected with an *ought*, or an *ought not*. This change is imperceptible; but is, however, of the last consequence. For as this ought, or ought not, expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it. But as authors do not commonly use this precaution, I shall presume

to recommend it to the readers; and am persuaded, that this small attention wou'd subvert all the vulgar systems of morality, and let us see, that the distinction of vice and virtue is not founded merely on the relations of objects, nor is perceiv'd by reason.⁴

Hume's point is somewhat obscured by the terms of his idiosyncratic philosophy, but the basic idea is simple: There's something illicit about using a series of entirely *non-normative* premises to establish a *normative* conclusion. The conclusion seems unrelated, or insufficiently related, to the premises adduced in support of it. Scholars ever since have generally agreed with Hume on this point, even though there's no consensus about exactly what is wrong with these arguments.⁵

And yet, in spite of this lack of consensus, we can say something illuminating about *is-to-ought* inferences, for they are eerily similar to question-begging arguments. Suppose Jeremy is considering telling his wife that he went bowling last night, and his friend Immanuel takes him aside and makes the following argument:

1. You were not bowling last night but having an affair with Rhonda.
 2. Premise 1 is something that you and I both know very well to be true.
 3. So in telling your wife that you were bowling last night you would be intentionally deceiving her.
-
4. Therefore, you should not tell your wife that you were bowling last night.

Never mind that we may at first be sympathetic to Immanuel's conclusion. A step seems to have been skipped between premise 3 and the conclusion in 4, and it's precisely that suppressed premise which Jeremy is likely to find objectionable. For, given that the first three premises are common knowledge between the two, Jeremy can't very well defend himself by denying the facts. So if he's going to disagree with Immanuel and deny that he shouldn't tell his wife he went bowling

⁴David Hume, *A Treatise of Human Nature* (1739), 3.1.1.

⁵The most common explanation one hears is that *is-to-ought* arguments are all logically invalid, but nothing could be further from the truth. Students of introductory logic can verify that "Tea-drinking is common in England" entails "Either tea-drinking is common in England or all New Zealanders ought to be shot," and that "All undertakers are church officials" entails "Undertakers ought to do whatever all church officials ought to do."

last night, he'll have to deny that 4 follows from 3. And he could do that by pointing out that Immanuel *would* be right *if* 3' were also true:

3'. One ought never knowingly deceive one's wife.

but then could go on to claim that 3' is false. After all, he and Rhonda are through: it was a one-night stand, a stupid mistake which in the end only made him better appreciate his marriage. Telling his wife the truth about last night would only damage their relationship and would do no one any good. So Jeremy believes 3*, which does not entail 4:

3*. One ought never knowingly deceive one's wife, except when failing to deceive would only do harm.

So the vice that Immanuel's argument commits is that it *obscures the debate*. The dispute is *really* a question of values: what is more important, honesty or a family's happiness? But Immanuel makes it seem as if it were as simple as a matter of fact — and, even worse, a matter of fact that Jeremy can't plausibly deny. This is good rhetoric, but it is not good philosophy.

Moreover, Immanuel's argument possesses this defect even if he is *right* and 3' is true. For just as with arguments that straightforwardly beg the question, the difficulty is not a matter of right and wrong but of whether *good reason* has been given to believe a certain conclusion.⁶ This diagnosis also explains when it's OK to jump straight from a matter of pure descriptive fact to a normative conclusion: namely, when the normative premise which has been omitted is not under dispute. Suppose, for instance, that I come up to you one day at work and make the following argument:

1. The building is on fire.

2. We should leave the building.

In this case it's OK to omit the premise "One should leave a building that is on fire" because duh.

⁶Note also that if 3' is true then, as with many question-begging arguments, Immanuel's will be valid. For if 3' is true at all then it's necessarily true. (Why? Give some thought to this. I intend "never" here to mean something like "under no circumstances whatsoever.") And if it's necessarily true then it's not possible for premises 1 through 3 to all be true and the conclusion 4 to be false.

So let this be a first lesson in how to read and write moral philosophy, in which normative claims are always what's under dispute: always explicitly state and justify the normative principles you are relying on to derive your conclusions, and beware of authors who try to pull the wool over your eyes by jumping straight from *is* to *ought*.

Doing Moral Philosophy: Reflective Equilibrium

So that's an example of how *not* to do moral philosophy. How, then, are you supposed to approach the texts you'll be reading and the papers you'll be writing in this class?

The remarks that I'll make here will of necessity be brief and hand-wavy, since philosophers don't exactly have an algorithm for finding moral truth (yet). But it is helpful to reflect on how people actually come to learn moral truths. You probably have a more nuanced moral worldview now than you did as a child, after all. Perhaps when you were a kid you saw some fictionalized murder on TV and were horrified, which prompted you to think, "All killing is wrong!" But maybe later you heard about some terrible killer's execution and you reflected that maybe not *all* killing is wrong, since people should be punished for crimes and sometimes that punishment should be capital. Maybe around that time in your life you also came to accept killing in war as justified. Finally, as you got older and started to think about the extraordinary litigation costs that capital sentences incur, and maybe also that giving people a second chance and rehabilitating them is more valuable than blind adherence to some kind of retributivist, eye-for-an-eye ethic, you reconsidered your reconsideration of the morality of capital punishment.

This kind of calibration of one's ethical view both by particular intuitions ("Gosh, what that killer did was horrible!") and by general moral principles ("We ought always give offenders a second chance") is called "reflective equilibrium." John Rawls, the giant of 20th century American political philosophy, characterizes it this way:

I now turn to the notion of reflective equilibrium. The need for this idea arises as follows. . . . In describing our sense of justice an allowance must be made for the likelihood that considered judgments

are no doubt subject to certain irregularities and distortions despite the fact that they are rendered under favorable circumstances. When a person is presented with an intuitively appealing account of his sense of justice (one, say, which embodies various reasonable and natural presumptions), he may well revise his judgments to conform to its principles even though the theory does not fit his existing judgments exactly. He is especially likely to do this if he can find an explanation for the deviations which undermines his confidence in his original judgments and if the conception presented yields a judgment which he finds he can now accept. From the standpoint of moral theory, the best account of a person's sense of justice is not the one which fits his judgments prior to his examining any conception of justice, but rather the one which matches his judgments in reflective equilibrium. As we have seen, this state is one reached after a person has weighed various proposed conceptions and he has either revised his judgments to accord with one of them or held fast to his initial convictions (and the corresponding conception).⁷

Reflective equilibrium isn't much of a recipe for doing moral philosophy, and if you read it uncharitably it doesn't sound like it's any more helpful than an admonition to "think really hard about what's good and bad." But that's only if you read it uncharitably. Here's a more concrete interpretation of the method:

- *Start with forceful, direct, particular, and deeply held beliefs — ones that you are unlikely to change no matter what you may later learn.* By "particular" and "deeply held" I mean beliefs like "Killing a passerby for the fun of it is very bad." I do *not* mean beliefs like "More liberty is always better" or "Lying is never permissible".
- *Hypothesize: What general moral principle could explain these deeply-held beliefs?* For instance, maybe killing passersby for fun is very bad because causing harm to anyone is very bad.
- *Interrogate this principle, and revise it if necessary.* Are there any cases in which, according to your deeply-held direct intuitions, it gets things wrong? For instance, you might think that it's not very bad at all to kill a mad scientist who is about to blow up the entire planet, even if that does cause

⁷John Rawls, *A Theory of Justice (Revised Edition)* (1971), pp. 42-3.

harm to someone. So maybe the principle we're looking for is "Causing *unnecessary* harm to anyone is very bad." This is a difficult step because it often involves looking for ways to disprove a principle you'd very much like to be true. Leaning on other people's moral intuitions at this step can be handy.

- *Draw conclusions from your investigations.* What follows from the general principle you now accept? This step may lead you to revise particular intuitions which you *thought* were forceful and deeply-held, but which you now see as unjustifiable. For instance, you may reflect that factory farming causes harm to animals, that this harm is unnecessary, and that animals are near enough to people to count morally. This, in conjunction with the general principle you now accept, may lead you to look at buying meat in the supermarket in a different light.
- *Return to step 1.* Since you have likely revised some of your intuitive reactions, you now should go back and see whether you can find other principles that would give coherence to your intuitions. Restarting this process may lead you to modify the general principle you accepted in step 2, but even if you do, you probably won't abandon it entirely. That's how reflective equilibrium leads us to make progress in our moral beliefs.

A Taxonomy of Normative Concepts: A Reference Sheet

Moral vs non-moral normativity

"Normativity" in the wide sense means any of the notions below. I tend to use the word in the wide sense. In the narrow sense it means either notions of *right* or the notion of a *reason*, depending on the context. ("Norma" means a carpenter's square in Latin; the idea is that of applying a rule. Reasons are often considered rule-like.)

Is there such a thing as specifically moral normativity? There's vast disagreement on this. Some think morality is all of normativity; some think it's just an appearance of normativity; most think there is moral and non-moral normativity. I'll presume the lattermost is true. That is, for instance, there are moral values *and* non-moral values.

What's the difference between the two? It's hard to say. Morality deals with the worth that people have just for being people, and what we owe to them just in virtue of their being people. (It might go broader than that: I have moral obligations to my mother in virtue of her being a parent.) The value of art isn't really like that; it doesn't seem to depend on the value of people or on relations between them.

Evaluative (axiological) concepts

Good, bad, value, disvalue; awesome, terrible, beautiful.

Values are graded (something can be more or less good), often can be weighed against each other, and do not by themselves entail any deontic notions. (Just because one action is good does not mean you should do it, since another action may be better.)

A very important subclass of evaluative concepts are the **aretaic** concepts, which deal with evaluating people's character: *honest, modest, kind, trustworthy, mean, judgmental, prideful, imprudent*, etc. These are all ways for someone to be good or bad as a person; they are virtues and vices.

Deontic notions

*Ought, should, duty, permissible, forbidden, right, wrong.*⁸

Deontic notions imply a threshold or standard. These too can generally be weighed against each other, as with competing obligations. The only exception is an all-out or all-things-considered norm, such as what you ought to do, all things considered, right now.

Moral responsibility

Blame, blameworthiness, guilt, what one is morally responsible for, what one can be held to account for doing.

These concepts cut across the evaluative/deontic distinction. Moral responsibility is different from, but related to, causal responsibility. You can be in part causally responsible for your own burglary (you forgot to lock the front door, and

⁸Sometimes evaluative terms are used in a deontic way, as when one says "The car's no good anymore since it doesn't go faster than 60" to mean "The car isn't good *enough* anymore."

a locked door would have deterred the burglar) but not be morally responsible (the burglar's to blame for the burglary, not you).

You can *do wrong* but not be blameworthy for it in case you have an *excuse*: for instance, a good excuse for robbing the bank is that Mobster Tony threatened to kill your family unless you did. Excuses are different from justifications, which show that you didn't do wrong after all. Killing in self-defense is generally considered a justification, not an excuse.

One is *guilty* of wrongdoing, or *culpable* for it, just in case one has done inexcusable wrong.

Reasons & rationality

In ethics these notions are connected to the faculty of practical reasoning, the psychological mechanisms by which we figure out what to do. The use of these terms in philosophy is highly contentious, and different writers mean different things by them. Below are the most common contemporary ways of using them, though in reading you should be wary of how authors might use them differently.

The notion of a reason is of a *fact that counts in favor* of some agent's performing some action. A reason for John to go to the party tonight is *that there will be dancing there*. Because he likes dancing, that's a fact that he should weigh in his deliberations about what to do. But most people think that *that there are people in the building* is a reason for John *not* to bomb it, whether he wants to bomb people or not.

A common notion of rationality has to do with *coherence*: it's irrational to think that you should leave the house now but also that you shouldn't. It's irrational to think you have most reason not to eat a third piece of cake tonight but decide to eat it anyway. Some philosophers, however, use "rational" to mean "responding appropriately to one's reasons".

Moral theories

Consequentialist theories These define what's *right* to do in terms of *the promotion* of certain *valuable states of affairs*. For instance, utilitarians believe that the right action to perform at any moment is the one that, among an agent's available options, maximizes the total amount of well-being in the

world. The state of affairs that's to be promoted is *people's having good lives*.

Deontological theories These deny consequentialism. Generally they don't think you can define what's right to do in terms of what's best to do. Instead there is, at the fundamental level, just a list of *duties* that need to be weighed against each other. Say, a duty to respect autonomy, a duty to be loyal and keep one's promises, and a duty to help others. These think of value differently than consequentialist theories. Deontologists will say that there's value to Jim's keeping his promise, but not a kind of value that means that anyone else should help Jim keep his promise. (When people write about "Kantian ethics", they mean a specific kind of deontological theory.)

Virtue theories These typically define what's right to do in terms of what a *virtuous* person would do in the circumstances. It starts with an idea of a *morally ideal person* and defines right in wrong with reference to that person's actions.