

13

The Theory–Theory Approach to Ethics

Frank Jackson

1. Methodological Preamble

As Wittgenstein noted, “naming something is like attaching a label to a thing”.¹ But what are labels good for? They help us interact successfully with the things we have labelled. It is not for nothing that we give the front door key a different label from the back door key. What is more, it is part and parcel of these successful interactions that our labels ground information-preserving causal chains. That’s how proper names help us arrive at the right cities when we travel. It explains, for example, why typing them into search engines or uttering them to travel agents produces the information we need. Reflections like these tell us that there has to be something essentially right about the causal theory of reference for proper names, though some (myself included) insist that we should understand it in its causal descriptivist guise.

Suppose we approached ethical terms in the same spirit, asking questions like: What do we do with them? What are they good for? How could we do something similar but do it better? This way of thinking changes the terms of the debate in ethics. Someone tells us, as it might be, that what’s morally right is that which an idealized version of ourselves would resolve to do given full information about the available options. The questions to ask become ones like: What purpose might be served by giving people this information? and, Why would it be good to establish that some course of action is the one an idealized version of ourselves would resolve to do?

This essay will explore the implications of this way of thinking about ethical terms, concepts, and properties, and explain why a theory–theory story about them is so attractive when we approach matters from this perspective.²

2. Background Presumptions

One has to start somewhere and it can help to make explicit one’s starting presumptions. Here are mine. (i) When someone affirms, “I believe that abortion is not always

¹ (Wittgenstein 1963: section 15).

² By the theory–theory I mean the view of, for example, Jackson (1992), Jackson and Pettit (1995); details are given below.

wrong”, this should be taken at face value, as stating what they believe. And their belief is true just if it is the case that abortion is not always wrong, and that’s the case just when the sentence “Abortion is not always wrong” is true. (ii) We marshal evidence for and against different views about what ought to be done. (iii) We change our minds about what we ought to do, sometimes as a result of the impact of new factual information (think, e.g., of those who have changed their minds about whether the water supply should be fluoridated) and sometimes as a result of a change in what’s valued at some fundamental level (think, e.g., of those who have moved from being average utilitarians to being total utilitarians). (iv) We debate issues like, Does an act’s being the morally best of those that are available entail that it is obligatory?

Taken together, considerations like these strongly support cognitivism in ethics, where by cognitivism I mean the view that ethical beliefs are beliefs properly speaking. For example, to believe that some course of action is morally wrong is to represent it as being a certain way, as having the property of being morally wrong, and the belief will be true just if it has the property in question. Moreover, a sentence like “Abortion is not always wrong” represents things as being a certain way, and will be true just when things are that way. Of course, strong support can be defeated, and some will offer one or another consideration to blunt the case for cognitivism. But I think that the overall case for cognitivism is more than strong enough to make it a reasonable starting point.

3. The Metaphysics of Ethical Properties

Cognitivism gives us license to talk about ethical properties in a metaphysically serious sense. Right acts differ from wrong acts in how they are in the same sort of way that tall people differ from short people in how they are, and the countryside in France differs from the countryside in England. Here we are not talking about properties in the sense of fundamental joints in nature or anything like that, and we are not talking about properties in the sense of universals.³ We are talking about properties in the sense that is in play when people wonder *how much* the housing market has dropped, or which antibiotic is the *most effective*, and, more generally, are wondering about what things are *like*, their nature.

Cognitivism also means that we know how to talk about ethical properties: use ethical terms. If we are English speakers, we will use “morally right”, “immoral”, “evil” and so on. But do we have to use ethical terms? Is it compulsory?

Many insist that we do not have to use ethical or moral terms. We have other terms we can use instead. I will call this view naturalism. Sometimes the case for it is made by reminding us of the plausibility of the view that the kinds of properties that figure in the natural sciences, broadly construed, give us a complete picture of the properties to be found instantiated in our world. But then each and every instantiated property can be picked out in the terms of the natural sciences, and those terms do not contain ethical vocabulary. This means that—provided that ethical properties are

³ To put matters in the terms of Armstrong (1978).

instantiated—each and every ethical property can be picked out without using ethical terms. Sometimes causal considerations are brought into the picture. We want ethical properties to figure in causal explanations, including explanations of our beliefs that one or another action possesses one or another moral property. How else could we know about them? And how else, given the plausibility of broadly causal approaches to content, could we have beliefs with the content that such and such an action is immoral, and that so and so a person is morally admirable, for examples? But we know enough about how our world works to know that the properties that figure in causal explanations are one and all properties we can pick out without recourse to ethical expressions. Sometimes supervenience enters the picture. The account of how things are given in ethical terms supervenes on the account given in non-ethical terms. How could this be the case, runs one argument in support of naturalism, unless the properties we pick out using ethical terms were the same properties (under different names, of course) that we pick out using non-ethical terms?

I have just stated naturalism (in ethics) in terms of language, as the view that we do not have to use ethical terms to pick out ethical properties, and reminded you of some familiar reasons for favouring naturalism, so understood. Why not put matters directly in terms of properties? For when you worry about what you ought to do, you aren't worrying about the words that apply to a proposed course of action. You are worrying about what properties it has (or so we cognitivists insist). True, if cognitivism is correct, the action will be what you ought to do just if "is what you ought to do" applies to it, or as we say it above, just if the property picked out by "is what you ought to do" is possessed by it. All the same, it is the property, and whether or not it is possessed, that concerns you, not the words *per se*. The wrongness or otherwise of abortion is not a question about words. The trouble with stating naturalism directly in terms of properties, is that one ends up saying something like: ethical properties are identical to non-ethical properties. This invites the thought that the view can be dismissed out of hand; its falsity comes from the meaning of the prefix "non-". It is, I think, safest to think of naturalism as holding that ethical properties can be picked out without using ethical expressions.

A view sometimes called soft naturalism⁴ holds that ethical properties are one and all natural properties, but that the ethical cannot be analysed in non-ethical terms. Have we somehow failed to make a place for soft naturalism? No. Soft naturalism isn't the view that right acts, for example, have a property we cannot pick out using non-ethical terms. When soft naturalists explain their view as holding that that something like

Being right = maximizing expected utility

is true, but that being right, or that "is right", or that the concept RIGHTNESS, etc., cannot be analysed in non-ethical terms and, in particular, cannot be analysed in terms of maximizing expected utility, they are not saying that being right requires ethical terms to pick it out. They cannot be saying that. If being right is identical with maximizing expected utility, Leibniz's Law tells us that there is a way to pick it out without using ethical expressions—use "maximizing expected utility". For "maximizing

⁴ To use the terminology of Parfit (2011: chapter 25).

expected utility” picks out maximizing expected utility (obviously), and if being right is identical with maximizing expected utility, then “maximizing expected utility” picks out being right. Of course, there is an important distinction between versions of naturalism that hold that ethical terms, or concepts or . . . can be analysed in terms of non-ethical ones, and versions of naturalism that deny that any such analyses are possible. But all versions agree—they have to agree—that we do not need ethical terms to pick out ethical properties.

Should we go along with naturalism? Mooreans—those who insist that it is compulsory to use ethical terms if one wishes to pick out ethical properties, on the grounds that ethical properties are quite distinct from the kinds of properties talked of in the natural sciences—have replies to each of the considerations outlined earlier. But I think, as do many, that, taken together, the considerations make a very strong case against Mooreanism and for naturalism. Moreover, there is much dissension within the Moorean camp. I don’t mean in the replies to the considerations sketched above (though there is dissension there also), I mean concerning the distribution of the ethical properties they insist cannot be picked out without using ethical terms. One should, I urge, be suspicious of claims to detect properties that are *sui generis* in the sense that they can only be picked out and explained in terms special to them, and whose causal explanatory roles with respect what happens in our world are, to say the least, unclear. But if those who claim to detect these special properties display an impressive degree of non-collusive agreement about the distribution of these special properties, sceptics have to sit up and take notice. It is, however, an only too familiar fact that there is great disagreement over the distribution of these allegedly special properties, especially over exactly which properties picked out in non-ethical terms ground one or another ethical property.⁵ For example, although Mooreans agree that, by and large, killing humans is wrong, as do nearly all of us, they differ greatly over exactly why it is wrong—over which property of killing humans picked out in non-ethical terms makes this the case—as the debate over eating meat makes vivid. The same goes for keeping promises, causing pain, allowing people to make free choices and all the rest. The general agreement that they are, by and large: right, wrong, and right, respectively, masks huge differences over why they earn these verdicts. Of course, dissension is not limited to Mooreans, but it is a special problem for them. When a group of theorists claim: one, to detect properties that are opaque to science; two, are unable to say much directly about their nature (that’s the cash value of saying that the properties are *sui generis*); three, grant that these special properties are grounded in ones that are transparent to science; while, four, being unable to agree among themselves about how the grounding works, surely scepticism about these properties is warranted.

4. Hunting for the Ethical after Moore

Turning one’s back on Moorean thoughts about ethical properties—that is, embracing naturalism as specified above—has a profound effect on how we should think

⁵ As Mackie (1977: 37) highlights.

about debates in ethics. This point has been noted but its full impact sometimes escapes explicit recognition, or so it seems to me. It can be very hard to think outside the Moorean box. The impact is that once one has a full inventory of all the properties we can pick out in non-ethical terms of some proposed course of action, or maybe all the properties anyone sensible could possibly think might be relevant, the question of whether or not the action is morally right is settled; there is nothing more to find out about the action. The perennially tempting thought that even when that full inventory is to hand, it is, in some good sense, a further question whether or not the action is morally right is a mistake.

Well it is not quite like this. The inclusive sense we are giving the notion of a property means that there is a sense in which one cannot have full inventory. As well as the properties picked out by, for example, “is in Australia” and “is causing pain”, there are different properties picked out by “is in Australia and causing pain”, “is in Australia or causing pain”, “is causing pain and contains an odd number of electrons”, and so on. There is, accordingly, no such thing as a full inventory in the sense of a very long but finite list that covers every property we can pick out in non-ethical terms. What we can have is a full inventory in the sense of a list of all the properties we can pick out in non-ethical terms that might plausibly be relevant (containing an odd number of electrons is an example of a property that is not plausibly relevant), along with a grasp of the various ways of combining those ingredients to form expressions that pick out further properties that might possibly be relevant, as happens when someone insists that what’s crucial for being a value is not being desired but being what’s desired to be desired, or someone else says that to be a right action is to be one which it is rational to desire everyone in similar circumstances perform, or some such.

When I say above that once one has a full inventory of all the properties we can pick out in non-ethical terms of some proposed course of action, the question of whether or not the action is morally right is settled, I do not mean by this that we should stop using words like “is morally right”, “is evil”, and so on. But the rationale for using them cannot be that they serve to pick out properties that have been left out of the inventory—there are no such properties to be picked out. The rationale for using them can only be that they pick out, from the properties already inventoried, properties that are best for some purpose or other, where “best”, in this context, is—*of course*—itself to be understood in non-moral terms. To fail to see this last point would be precisely to slip back into the Moorean camp. But now a key question becomes how to understand this sense of “best”. What does it take for some selection of properties we pick out using non-ethical terms to be the best ones to pick out using ethical terms?

5. Making a Start on What It Takes to be the Best Properties for Ethical Terms to Pick Out

Here is a simple example to set the scene. Physicists use the term “the centre of mass” to pick out an important property of a system of particles. There is a sense in which the term is not essential. They could have picked out the very same property using the

terms for mass and position. Although having such and such a centre of mass is distinct both from having so and so a mass, and having this or that position, the terms for mass and position, suitably deployed, pick out the very same property as having such and such a centre of mass. That's what gets done when the centre of mass of a system of particles is specified in a physics text. But there is, of course, no mystery about the rationale for having the term "centre of mass" in physics, the value of the concept, and the importance of the property picked out: centres of mass play important predictive and explanatory roles.

What makes this example an easy one to think about (which is, of course, why I chose it) is that we know the relevant criterion for being a property that it would be good to pick out in physics and, thereby, a property it would be good to have a term for—it is being good for doing what physicists' do. Information about a system of particles' centre of mass is useful for predicting and explaining the system's behaviour over time, and predicting and explaining the behaviour of systems of particles over time is among what physicists do. What is far less clear is what it takes to be a good property to pick out in the ethics case. We know what physics is for, and that guides us in evaluating, say, $1/2Mv^2$ versus $1/2M^2v$, and explains why we give a tick to the first (when v is not too near c) and a cross to the second, and why we give a tick to centres of mass. But what is ethics for? That's the question we need an answer to if we are to find out which properties ethical terms pick out.

6. The Problem of Finding Purpose in Traditional Debates in Ethics

What ethics is for is a question about its purpose. But much of the debate in ethics does not look like a search for its purpose. It looks instead like a complex exercise in intuition swapping, combined with attempts to shift one's opponents' intuitions by noting one or another tension in their intuitive responses to different cases. Indeed, it does not just look like this; that's what much of it is. And some have, understandably, expressed concerns about this methodology. Do we do serious science by intuition swapping? Now, in fact and of course, intuitions do play a role in science and in theory building more generally. We perforce have to start our theorizing somewhere, and, naturally and properly, start from that which we find most intuitively appealing. The worry is the extent to which debate in ethics is driven by intuitions. A huge amount of it consists in painting pictures of possible situations using written and spoken words, and urging that one or another intuitively appealing response is indeed the correct response to one or another described situation. Sometimes the protagonists describe themselves as doing conceptual analysis; sometimes the exercise is given a more empirical flavour, as happens in the work of those who describe themselves as experimental philosophers. But it is intuition swapping all the same. The difference is that those doing conceptual analysis are more likely to insist that certain intuitions—their own, after due reflection—are the only defensible ones, whereas those doing experimental philosophy are more likely to want to collect intuitions from a range of subjects. But both are harvesting intuitions from word pictures, and building their theories on what gets harvested.

Is there a way of understanding the process I have just given a rough description of that makes it defensible and, moreover, fits with our leading idea that we should look for the properties it would be good to use ethical terms for—the idea we introduced in the first section by noting how a causal theory of reference for proper names drops out of asking what proper names are good for? Perhaps surprisingly, the answer to this question is yes, and the way to see this is to reflect on what we learn from a famous argument for analytical functionalism in the philosophy of mind, or so I will argue.

7. Analytical Functionalism

Analytical functionalism is a view about the meanings of certain words, words like “belief”, “desire”, and “feeling hungry”. The view is that they are susceptible to an analysis in functional terms. But combined with this view about mental language is a position on the metaphysics of mental states. Sometimes the claim concerning the metaphysics of mental states is that they are functional states; sometimes the claim is that they are states (brain states, in fact) that get to be the mental states they are by virtue of the functional roles they play. The difference does not matter for our concerns. What matters for us is how the two positions, one about words and the other about the nature of mental states, are related. And here’s a plausible answer to that question, or so it seems to me and many. When we describe human beings (and other animals) using mental state terms—“belief”, “desire”, and all the rest—something remarkable happens. Our ability to make successful predictions and explanations concerning their interactions with the environment and each other goes up dramatically. This strongly suggests that in using mental terms to describe humans, we are picking out properties humans in fact have. That’s the obvious explanation of the utility of using these words. What properties might they be? Well, what’s remarkable is, as we noted, the increase in predictive and explanatory power concerning certain interactions, and we know that functional properties are especially valuable for predicting and explaining interactions. This leads us to the hypothesis that mental words pick out functional properties. This leaves us with a choice between saying that mental states are the functional properties themselves, or are states having the functional properties in question. Either way, we have an explanation of why mental language is so useful, and one that connects facts about words with the metaphysics of mind.

The purpose of this reminder is to nail down the point that what happens when we use certain words can tell us a lot about the nature of that which we are using those words for, and that’s something you can sign up to even if you dissent from one or another detail in the little example given above.

I hope the next question on my agenda will now be obvious. What do we learn from our use of ethical words, and how does this relate to the intuition swapping methodology of so much of the debate in ethics?

8. The Effects of Using Ethical Words

What happens when we describe matters in ethical terms? When dealing with psychopaths, nothing much. But when dealing with most of us, something very

striking happens. Certain kinds of disputes get settled, or more nearly settled, when we use ethical terms. Suppose that Crusoe was alone on that island; in that case, he can do what he wants to do. He does not have to balance what he wants against what others want. But for most of us, a recurring issue is how to balance what I want against what you want. I want to see the doctor first, so do you. We cannot both get what we want. I want the roads near my house to be of good quality, you want the roads near your house to be of good quality. But there's only so much funding to go around; we cannot both get what we want. And so it goes. The fact that we live in communities and that resources are limited means that we cannot get exactly what we want, and sometimes cannot even get close to what we want. Fighting is one solution but its drawbacks are notorious. The solution that is often adopted is to describe matters in ethical terms. Sentences like "Your illness is more serious than mine (or you were here first), so the doctor ought to see you first" and "The road past my house services many more people than does the road past your house, so the bulk of the funding should go to it" start to get uttered. And when they do, the dust settles. I take it to be a commonplace that using ethical terms can help when dealing with clashes in wants. The exact story about the how and why of this might take many forms, but what matters for our purposes is that it happens. Framing issues in ethical terms helps settle hard decisions about the distributions of goods in the sense of the things we desire. I emphasize that many have said things like this. Here is a recent statement (Sterelny and Fraser 2017: abstract) "moral facts are facts about co-operation, and the conditions and practices that support or undermine it". Our concern is with the fact that using moral terms helps resolve conflicts; their concern is with moral facts and the right evolutionary story concerning them. But the song is the much same nevertheless. The words could hardly be so useful unless they picked out properties in the world, and the instantiation of these properties are the facts Sterelny and Fraser are talking about.

Although I am prescinding from the details of how using ethical terms helps resolve disputes and aids co-operation, we can say this much in the broad. Part of the process involves motivation. When matters get described in moral terms, it affects people's motivations. If that were not true, there would be no reason to use moral terms. When someone uses "is what ought to be done", or the like, of some proposed course of action, it is no accident that they tend to bring that course of action about. I want to see the doctor first. You want to see the doctor first. I learn the details of your situation, and come to a conclusion that I express in the words, "You ought to see the doctor first". I make way for you, and you feel good about my doing so. And so it goes.

We can now see how to make sense of what's going on when we swap all those intuitions, and do so in a way that makes the process a reasonable one. We are swapping intuitions about word pictures, as we say above. Why should we pay attention to these intuitions? Well, the words are our words; the intuitions are our first up opinions about the application of those words to those situations. (We don't use an "intuition meter"; we simply say what we believe about those situations using the words in question.) But our words would not be much use for describing the world if our opinions about when they do or do not apply had no weight. So, to turn to the ethics case, our intuitions are a guide to what properties our ethical terms pick

out. What's more, as we have lately been outlining, we have good reason to hold that the properties our ethical terms pick out are good for resolving the kinds of conflicts that are part and parcel of being creatures that need to resolve conflicts in interests and, more generally, co-operate as members of communities. A ditty runs "Things go better with coke"⁶. Communities go better when issues get framed in ethical terms and the results of those framings are implementations of the actions that got described as ethically right.

9. How Ethical Terms Can Get to Make Things Go Better: A Bit More Detail

Using ethical terms makes things go better, especially in situations where a group of people cannot each get what they separately want, and in cases where entering co-operative arrangements have clear benefits. That's what we have just been saying. Let's now spell things out a bit more.

The way we use ethical terms can be divided into three categories.⁶ One consists of connections between matters described in non-ethical terms and matters described in ethical terms. Killing is typically wrong. One ought as a rule to keep one's promises. If some action is impossible, it isn't something that ought to be done. Pain is bad. And so on. I will call these clauses input clauses. They get to be called input clauses because they can be framed as conditionals with non-ethical antecedents and ethical consequents. For example, the last illustration can be rephrased as: if *x* is pain, then *x* is bad; and of course the penultimate example is already in the form in question. If using ethical language is to make things go better, input clauses are vital. We need them to provide ethical discourse with a good part of its subject matter—the actions, motives, policies, etc., that are up for evaluation in ethical terms. We talk above of the intuition swapping that is so prominent a feature of debates in ethics. Much of the swapping concerns input clauses. The debate engendered by the famous trolley car problem is one illustration. That debate is essentially a parade of variations on the theme of killing one to save many, accompanied by invitations to assign one or another verdict framed in terms of what ought to be done to the variations, all in the hope of finding a consistent pattern in the verdicts, or, better, a consistent pattern in the considered verdicts.

A second category consists of inter-connections between matters described in ethical terms. Here are some rough examples. Rights imply duties to protect those rights. What ought to be done is what's best out of the available options. You ought to do whatever is a necessary condition for what you ought to do. I call these clauses internal clauses. They are a vital part of the way we reason about connections between matters described in ethical terms. A simple example is what happens when we move from matters described in terms of what is good or bad to what we ought to do, maybe balancing the extent to which we have special moral obligations towards those close to us. Internal clauses provoke their share of the intuition swapping that dominates debates in ethics. Think of the intuition swapping that

⁶ What follows draws on, for example, Jackson (1992 and 1998).

goes on in discussions over how to balance obligations to those we are close to against a general duty of beneficence. Or, again, think of the intuition swapping that goes on in discussions of whether or not an obligation to do A and B implies an obligation to do A, and in discussions of how being morally permissible relates to being morally obligatory.

Finally, there are output clauses. They connect people's beliefs framed in ethical terms with what they are motivated to do. They are every bit as much a subject of intuition swapping as are the input and internal clauses. Some find it intuitively compelling that someone who believes that they ought to do A will be motivated to some extent to do A. Dissenters present counter-examples—subjects who agree that A ought to be done but show no inclination whatever to do A. Dissenters to the dissenters respond by urging that these cases are one's of subjects who are only agreeing in "words" that A ought to be done; they haven't taken on board what believing that A ought to be done really amounts to. But, however this debate pans out, it had better be the case that, very often, describing matters in ethical terms as makes a difference to behaviour, as we noted in an earlier section.

I hope that what I have just been saying sounds pretty commonsensical. How could using ethical terms have good effects if there were no agreement about which actions, policies, characters, etc., are apt for description in ethical terms, no agreement about how to reason using ethical terms, and no agreement about what often happens when someone uses one or another ethical term to describe an action, policy, etc.? We need those input, internal, and output clauses. But insisting that there must be some agreement is not the same as insisting that there must be great agreement, and there isn't. It is an only too familiar fact that the intuition swapping we have mentioned a number of times has not led to widespread agreement on a core set of precisely specified input, internal, and output clauses. There is a reasonable amount of agreement in the rough, but not once we seek precise statements of the clauses.

10. How to Respond to the Disagreement

Someone may say, "So what", when asked about the disagreement. Perhaps they say that they have just published a book in ethics and in that book they nail down exactly why it is typically wrong to kill, and exactly what the exception clauses look like, and likewise for the other input clauses. They also insist that they capture the logic of ethics in the sense of the logical relations that hold between being right and being good, between being morally permissible and being right or wrong, between an obligation to do A and B and an obligation to do A, and so on. There is also a chapter on the connection between moral judgement and motivation, where, our author claims, that perennial issue is sorted once and for all. Perhaps they finish up by saying that they are sorry that there is so much dissension among theorists in ethics but hope that their book will reduce this and, anyway, when was philosophy an exercise in counting heads?

But how did our author make their case in their book? With experiments? Not in the sense of experiments in science. Finding the one true set of input clauses, for example, is not like finding that gold resists corrosion, or that light is a first signal. And the same goes for the internal and output clauses. With experiments in the sense

experimental philosophers sometimes have in mind—that is, by garnering responses to vignettes and reflecting on those responses? But that’s a version of intuition swapping, and the failure of this method to deliver consensus is exactly what has led to the discussion we are now having. By finding a new way to probe the true nature of the special properties that are picked out by ethical terms? No; to say that is to forget the implication of turning one’s back on Mooreanism. There are no such special properties. By appeal to the causal theory of reference, arguing that careful investigation of the causal origins of our use of ethical terms will reveal their referents and allow us to read off from these discoveries the one true set of (precisely specified) input, internal, and output clauses? But the right theory of reference for ethical terms has to be one or another kind of description theory, where by a description theory for a word “W”, I mean a theory that says that “W” applies to x just if x has a certain nature, that is, instantiates a certain property.⁷ Why do I say this? Because of the point we have adverted to already. People who worry about the morality of abortion are not worrying about words. They are worrying about the kind of action abortion is; they are worrying, that is, about abortion’s nature, whether or not it has a certain property. We perforce use words to discuss ethical questions, but these words need to pick out properties if our words are to engage with what matters ethically. I know non-cognitivists will object, but we cognitivists insist that the job of ethical terms is to describe, and that our ability to address what’s really at issue in debates in ethics depends on this fact. When we produce sentences like “Abortion is sometimes morally right” and invite discussion, a necessary condition for what follows to be relevant is that “is sometimes morally right” applies to abortion if and only if it has the very property people care about when they debate the morality of abortion.

How then might our imagined author make their case? And this is, of course, a pressing question for real authors of works in ethics. The intuition-driven nature of so much of the debate in ethics, combined with the diversity in intuitions, raises a serious question as to how any author might sensibly hold that they’d got it right. I think that there are only two ways our author can respond.

The first is to remind us of the hope many of us had when we first heard about the trolley problem. It is the hope that drives players of what has come to be called “Trolleyology”. It is the hope of finding a set of consistent and compelling responses to the original problem and its many variants. By consistent, I do not mean logically consistent. I mean that the set no-where contains different responses to, let’s say, version 3 and version 5 of the problem, where the only difference between the two versions is a property that no-one thinks is morally relevant. By compelling, I mean that the set is such that anyone who understands the various versions, and does not themselves fall into inconsistency in the sense just explained, will come to share the responses. So, the first way for our author to respond is to urge that when people reflect on what their book has to say, something remarkable will happen. There will be a massive reduction in the extent to which intuitions about ethics’ famous problem cases diverge, and that, although there will be holdouts (that’s how it is in

⁷ In *this* sense, a description theory of reference is true for the word “circle” in English. It applies to x just if x has a certain geometric property.

philosophy), it will be plausible that the holdouts are in some way confused or in some sense mean something different from what most of us mean when we use ethical terms.

Some will say, “Good luck with that”. I am not that pessimistic. As someone sympathetic to consequentialism, I have always taken the following line of argument seriously. Make the hypothesis that being right = maximizing expected utility, and see what ensues. What ensues, as we all know, are a range of verdicts about what’s right that clash with commonsense morality. Or at least they seem to, but, consequentialists urge, on examination matters are not so clear-cut. That’s the burden of, for example, Smart (1961) and Kagan (1982). So one way to go in response to the divergence in intuitions is to urge that, after due reflection, the detection of confused thinking and so on, the divergence will disappear.⁸ But, nowadays, my money is more on the second way to go, and here I am influenced by Alexander (1891–2), the paper by Sterelny and Fraser mentioned earlier, and, most importantly, by the examples of proper names and mental state terms.

We noted that it is matter of record that describing matters in ethical terms makes things go better, especially when dealing with problems that arise from the fact that we live in communities and need to adjudicate between competing claims. It is very hard to believe that this is an accident, as we noted. But if it isn’t an accident, there must be a story to tell about the properties our ethical terms are picking out, which explains this happy result. The situation is akin to that with names and mental state terms. It is a matter of record that the way we use names helps us find our way to conferences, etc. It is a matter of record that the way we use mental state terms helps us explain and predict what people will do. Both facts call for explanation. In first case, the explanation will advert to some version or other of the causal theory of reference for names.⁹ In the second, the explanation will advert to some version or other of functionalism, or so I suggest.

What’s the right story in the case of ethical terms? We can say this much, based on our earlier remarks. The story will have input clauses, internal clauses, and output clauses. We saw above that they are essential if moral terms are to be of use to us in negotiating our interactions with others in our communities. The task then is to find the best candidates to be the properties picked out by our moral terms to make true the input clauses, internal clauses, and output clauses in a way that best explains how they—our use of moral terms by virtue of the properties they pick out—make things go better in our interactions with our fellows. But, I hasten to add, properties first identified via their role in facilitating relatively amicable negotiations within a community can and will be possessed outside that context. Nothing I say here implies, for example, that we lack moral obligations towards those outside our own community.

Finally, this way of looking at things does not preclude the consequentialist’s answer. Maybe taking “is right” to pick out maximizing expected utility, along with

⁸ That is, the identity of being right with maximizing expected utility will drop out of *mature* folk morality, in the terms of Jackson (1998: 133).

⁹ In my view (as I said earlier), the version known as causal descriptivism. See Kroon (1987) and for the particular version I favour Jackson (2010: lecture five).

the corresponding choices for what “is morally good”, etc., pick out, deliver the best explanation. But maybe what does the best job is taking “is right” to pick out that which an idealized, fully rational version of ourselves would resolve to do given full information about the available options, along with the corresponding choices for what “is morally good”, etc., pick out.¹⁰ Or maybe what does the best job is taking “is wrong” to pick out actions which violate principles for the regulation of behaviour that no one could reasonably reject as a basis for informed, unforced, general agreement, along with the corresponding choices for “is permissible” etc.¹¹ Or maybe... I deliberately leave this key question open.

Acknowledgments

Thanks to the many who have discussed ethics with me over the years, with special thanks to David Braddon-Mitchell, Philip Pettit, David Plunkett, Peter Railton, and Michael Smith.

References

- Alexander, S. 1891–2. Is the Distinction between ‘Is’ and ‘Ought’ Ultimate and Irreducible? *Proceedings of the Aristotelian Society* 2 (1):100–7.
- Armstrong, D. M. 1978. *A Theory of Universals*. Cambridge: Cambridge University Press.
- Jackson, Frank. 1992. Critical Notice of Susan Hurley: *Natural Reasons*. *Australasian Journal of Philosophy* 70 (4):475–87.
- Jackson, Frank. 1998. *From Metaphysics to Ethics*. Oxford: Oxford University Press.
- Jackson, Frank. 2010. *Language, Names, and Information*. Oxford: Wiley-Blackwell.
- Jackson, Frank, and Pettit, Philip. 1995. Moral Functionalism and Moral Motivation. *Philosophical Quarterly* 45:20–40.
- Kagan, Shelly. 1982. *The Limits of Morality*. Oxford: Oxford University Press.
- Kroon, Fred. 1987. Causal Descriptivism. *Australasian Journal of Philosophy* 65 (1):1–17.
- Makie, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin.
- Parfit, Derek. 2011. *On What Matters*, vols 1 and 2. Oxford: Oxford University Press.
- Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Smart, J. J. C. 1961. *An Outline of a System of Utilitarian Ethics*. Melbourne: Melbourne University Press.
- Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.
- Sterelny, K. and Fraser, B. 2017. Evolution and Moral Realism. *British Journal for the Philosophy of Science* 68 (4): 981–1006.
- Wittgenstein, L. 1963. *Philosophical Investigations* (2nd edn). Oxford: Blackwell.

¹⁰ To draw on Smith (1994).

¹¹ To draw on Scanlon (1998).