Norms, Natures and God

Alexander R. Pruss

Contents

Acknowledgments

Central ideas for this paper were developed as part of the Wilde Lectures in Natural and Comparative Religion at Oxford University, Trinity Term, 2019.

Acknowledgments

I would like that thank ... Nicholas Breiner ....??

CHAPTER I

Introduction

1. Introductory remarks

I have a human nature or human form that governs my voluntary and involuntary activity. Much as the government governs the activity of the people *both* by legislating norms and encouraging people to follow the norms, my nature's governance also has the dual role of setting norms for me and influencing my activity to follow these norms. This nature is something real and intrinsic to me, something that makes me be what I am, a human being.

When extended to other fundamental beings besides humans, the above is the center of Aristotle's metaphysics. I will show that this center is extremely fruitful, providing compelling solutions to problems in ethics, epistemology, the philosophy of mind, semantics, metaphysics and philosophy of science. Many of these are prominent problems that have been the subject of much discussion, such as the problem of priors in Bayesian epistemology or of vagueness in semantics, while others are problems that have not attracted much attention, such as the problem of seemingly arbitrary detail in moral rules. I shall discuss these solutions in Chapters II–IX.

The ability to give unified solutions to an array of problems spread through many areas of philosophy gives one a very good reason to accept the central Aristotelian theses. However, in Chapter X, I will also argue that this center cannot hold on its own, and the way to be an intellectually satisfied Aristotelian, especially after Darwin, is to be a theist as well.

There are several lines of thought readers attracted to the unified Aristotelian solutions may follow. Some may deny that the problems facing the central Aristotelian theses are as

serious as I contend. Some may agree that the problems are serious, and regretfully reject the Aristotelian apparatus, either because they take the cost of the theistic solution to be too great or are unconvinced that the theistic solution works on its own terms. Others may agree that the problems are serious but find some other solution than the theistic one. But some, I hope, will conclude that the Aristotelian solutions are so attractive, and the theistic solution to the problems is sufficiently plausible, that this book provides not only a good reason to accept the Aristotelian center but also to accept theism.

We will be elaborating the metaphysical apparatus of what I have been calling the "Aristotelian center" gradually?? as we move through the problems and details of their solutions. At the same time, not every detail of the solutions needs to be adopted by the reader to find the general Aristotelian strategy compelling. Finally, in Chapter XII we will collect together the needed aspects of the Aristotelian metaphysics and discuss in greater detail the metaphysics needed.

??paths through the book?

In the rest of this chapter, we will do two things. First, I will sketch the central Aristotelian metaphysics in slightly greater detail. Second, I will discuss a neglected science-based argument from the 17th century polymath Marin Mersenne for the existence of God. This argument does not work, I will argue. However, an important thread running through this book will be how "Mersenne problems" analogous to the problems in science raised by Mersenne arise in many areas of philosophy and provide a compelling case for the existence of Aristotelian natures or forms.

2. Aristotelian natures

2.1. A quick introduction. According to Aristotle, reality is fundamentally built out of substances, which are real mind-independent entities. These substances are not limited

to microphysical entities like quarks and photons—indeed, it is not even clear that the microphysical entities are substances at all[1]—and indeed Aristotle takes biological organisms like oak trees and human beings to be paradigm cases of substances.??ref

Each material substance has a form or nature—I will use the terms interchangeably in this book. This form or nature performs a number of roles including unifying the matter of the substance into a single thing, setting norms for the structure and activity of the substance, and guiding the actual development and activity of the object. The nature of the oak tree is not merely an arrangement of its particles, since an arrangement lacks normative force. In living things, the form of the substance is its life or soul: it makes the substance be alive.

Natures are innate to their substances. Nonetheless, this statement underdetermines an important question, namely whether substances of the same sort—say, red oaks—all numerically share one nature or each individual substance has its own nature, albeit in relevant respects??forwardref they are all exactly alike in substances of the same kind. For two things could in principle share something innate to them. It could be that all people have the same soul, much as two conjoined twins could have the same stomach. Aristotle scholarship is divided on the question whether Aristotle believed in "individual forms", one per substance. However, at least one of the advantages of an Aristotelian theory of form will be accentuated if we accept individual forms, as we shall see.??forward Further, there is good philosophical reason to take natures to be individual, as we shall see in ??forward. Thus, I shall take natures to be individual. Nonetheless, if you like shared forms, *many* of the benefits I will draw out for a theory of forms will be ones you, too, can have.

2.2. Aristotelian optimism. Natures not only define how a thing should function, but also actively lead the thing to function in that way. This means there is an inherent bias in each substance towards acting well. This bias leads to Aristotle's optimistic thought that

[1]The fact that in quantum mechanics, one can have a superposition of states with different numbers of particles is evidence that particles are not substances.??

natural states occur "for the most part"??ref, which is quite useful for figuring out what is in fact natural, since the frequency of the occurrence of a state is evidence of its naturalness.

There is, however, a tension in Aristotle's own thought between the above optimism and the pessimistic observation that most human beings are morally bad.??ref Aristotle may be empirically wrong about most people being bad??refs?, but nonetheless exploring the tension will help us understand Aristotelian optimism more clearly as it faces the problem of moral evil.

There are many substances with different natures in the world. The flourishing of some requires involves the languishing of others: the lion's feeding is the gazelle's death. Moreover, a substance's nature directs it to behavior that works well for the substance in its natural niche. But things do not always stay in their niche. Because of this, Aristotle has many resources for explaining why there is a significant set of cases where substances find themselves in unnatural states.[2] But Aristotle nonetheless thinks that misfortune will only be a minority of the cases.

Let us return to the Aristotelian optimism that things function well "for the most part". What is and is not "for the most part" depends on the reference class. Most humans have legs, but most living substances do not. If the reference class of the "for the most part" is all activities of all substances, then human moral behavior forms such a small portion of that class—it is so outnumbered by bacterial reproduction, say—that even if all human moral behavior were wicked it would be unlikely to make a difference with respect to the Aristotelian optimism. However, at the same time, with such a broad reference class, the optimism would be of little use to us in understanding normativity for humans, for humans could simply be an outlier in all respects.

A more optimistic reference class would be all the activities of a particular kind of substance. On this reading, Aristotle would lead us to expect that each kind of substance does well in most of its activities. But moral activity is only a small proportion of the activity of a human. We also breathe, we circulate blood, we repair cells, etc. Leibniz

[2]For further discussion of the harmony between substances, see ??forwardref.

estimated that three quarters??check,ref of our activity is at an animal level. Stalin was a complete moral failure, but still he maintained homeostasis until the age of 74. Human moral activity could, thus, be mostly bad even though most human activity is good. Again, the tension between Aristotle's general optimism and his pessimism about human morals would be resolved.

A yet more optimistic reference class would be a particular major type of activity—say, moral activity or reproduction—of a particular kind of substance. Now we would have the prediction that most human moral activity will be good, and this seems to contradict Aristotle's thesis about typical human moral badness. But even this is not clear. In MacDonald's *The Princess and Curdie*, Curdie has just expressed to the princess's grandmother a pessimistic thesis that unavoidably most things humans do are bad.

> 'There you are much mistaken,' said the old quavering voice. 'How little you must have thought! Why, you don't seem even to know the good of the things you are constantly doing. Now don't mistake me. I don't mean you are good for doing them. It is a good thing to eat your breakfast, but you don't fancy it's very good of you to do it. The thing is good, not you.'??ref

The old woman makes two important points. First, we should not forget that we perform *many* morally significant actions each day. Curdie ate breakfast. He could have thrown it at his mother, or just ungratefully poured it out on the grown. His eating breakfast was morally good. And we perform many such morally good actions each day. Second, the fact that we perform these morally good actions does not do us much credit, the grandmother insists. I suspect that the reason for her pessimism here is Curdie's lack of the kinds of motivations that would render breakfast-eating positively creditable. But the mere motivation to nourish himself was already good, even if not particularly creditable.

There is a further point we may add. While on a mathematics exam, it might be enough to get 60% to pass, morally speaking it is not enough that 60% of one's actions be good.

If in the morning I kick a neighbor's puppy, at lunch I charge my private meal to a research budget, in the afternoon I plagiarize something from a foreign language journal for inclusion in my book, and in the evening I cheat in order to beat my kid at chess, I am a bad person even if each of these actions is paired with two morally good actions of the eating-breakfast level of goodness. Having a majority of one's actions be good is not nearly enough to avoid being bad.

Thus with the reference class of "for the most part" restricted to moral activities, Aristotle's optimism and pessimism can be both maintained. And the above considerations also show that Aristotle's optimism is quite compatible with realism or pessmism about human morality.

A further optimistic ingredient that we will at times draw on is the idea that the different ways of being well in an organism have a tendency to mutually support each other in a unified kind of way. There will be trade-offs, sometimes tragic ones, but by and large a healthy heart supports healthy lungs, a healthy mind supports a healthy body, courage supports justice, justice supports courtesy, and courtesy supports kindness, all of which tend to make one live a happier life even by hedonistic standards.

3. Mersenne questions

3.1. Mersenne's argument. Marin Mersenne was a monk, philosopher, theologian and the 17th century equivalent of the arXiv preprint archive—he was a crucial line of communication between a broad variety of thinkers and scientists. He drew on his broad knowledge of the science of the time to offer an argument that begins with many pages of questions, of which the following are representative:

> Who gave more strength to the lion than to the ant? Who made it be that earth is not in the moon's place, and that the planets aren't larger or smaller, closer or further? Who has ordered all the parts of the world as we see them? ... Why is the moon 56 earth-radii away from the earth? Why is the sun 1182 [earth-radii] away from us at its apogee? ... and

> why is its distance at perigee not other than 1101 [earth-radii]? ... I could equally ask you about Saturn, and Jupiter, and Mars ...??refs[3]

These "Mersenne questions" go on and on, with a mind-numbing number of examples. And Mersenne has one answer to all these questions, posed in a rhetorical question: "Was it not God?"??ref

The argument sounds similar to fine-tuning arguments for theism which became popular in the late 20th century. These arguments, too, list a variety of physical parameters and offer God as an explanation of them all.??ref But there is a crucial ingredient that the fine-tuning arguments, namely that the parameters listed are needed for intelligent life as we know it, or for some other valuable trait of the universe, like its amenability to scientific investigation.??ref The basic idea behind the fine-tuning argument is, very roughly, that nature is indifferent to value but God cares about value, so the fact that the parameters are valuable provides evidence for theism over naturalism.

It is, thus, natural to look in Mersenne for arguments that it is particularly valuable for the moon to be 56 earth-radii from the earth, but at least in this work, Mersenne does not supply them or even hint at them. Nor is there any argument that it is better that lions are stronger than ants, or that it is better for the moon to orbit the earth rather than the other way around. If Mersenne is giving a fine-tuning argument, the argument is oddly incomplete. And Mersenne's penchant for adumbrating detail at great length makes it unlikely that he has simply omitted such a crucial part of the argument.

Rather, it appears that Mersenne is simply looking for an explanation of the scientific details he cites, sees no prospect of a scientific explanation, and offers theism as the alternative. And indeed it is only in the 20th century with computer models of solar system formation that we have much in the way of plausible answers to Mersenne's questions about the distances between solar system bodies. For instance, the leading theory of lunar formation involves the earth being hit by another body and a large chunk being pushed

[3]The moon-earth distance is approximately correct. The earth-sun distance is an order of magnitude off.

into orbit. Given assumptions about the impact, one can then explain the resulting distance between the earth and the moon. But notice that such an explanation only gives an answers to the Mersenne question about the earth-moon distance at the cost of raising similar Mersenne questions about the parameters of the impact such as the mass distribution of the pre-impact earth, the angle and location of impact, the mass distribution of the impacting body, etc.

But Mersenne has a fatal argumentative flaw. Even if we grant that it is very unlikely that a future science will predict these exact numbers, there is always the possibility of a stochastic explanation, one that does not predict exact values, but supposes a random natural process that generates a set of values at random. Now, if Mersenne had an argument showing that the values of the parameters are suspiciously valuable—say, necessary for intelligent life—then a stochastic explanation might not be as good as a theistic one. From a Bayesian point of view, we might be able to argue that it is extremely unlikely that a random selection of parameters would have such value, but not nearly so unlikely that God would choose such parameters and hence the data supports theism over randomness. But given that Mersenne makes no case that the parameters have anything to recommend them to God for creation, we have no reason to think that the probability of God choosing is these parameters is any higher than the probability of them arising randomly, and hence we have no support for theism.

Suppose, however, that we had a Mersenne-type case where randomness was not a satisfactory explanation. Then there would still be one more problem with the argument. If one is willing to deny the Principle of Sufficient Reason, one could simply say that the parameters are what they are and there is no reason why they are like this—that they are a *brute* fact. This, however, is less satisfying than the stochastic answer, for adverting to brute fact should be a last resort, to be chosen when no explanation is available. But here there is an option, namely theism.

3.2. Appearance of contingency. Mersenne gives a dizzying number of examples, and he seems to relish the sheer appearance of arbitrariness of the numbers like “56” and

"1182". While this has some rhetorical force, it also has argumentative force. The more arbitrary-looking parameters the parameters are, the less epistemically likely it is that they are what they hold of necessity or that good scientific theories will predict their exact values. And the greater the number of parameters, the less likely it is that science can provide an explanation of them all.

The appearance of arbitrariness is evidence of contingency, and contingency calls out for explanation.[4] But at the same time, we have to be careful here. For instance, it might seem arbitrary that protons have (approximately) 1836 times the mass of electrons, but the massses of protons and electrons could well be essential properties of them, so that a pair of particles whose mass ratio were different from 1836 could not be a proton-electron pair. So in some cases, the arbitrary-seeming parameter does in fact hold of necessity. But that does not mean that the Mersenne question disappears. For while the parameter itself is not contingent in these cases, there is contingency "nearby". Even if the masses of protons and electrons are essential properties, it is possible to have particles with similar behavior but other masses, and it will be contingent that the world contains a pair of opposite-charge particles with mass ratio (approximately) 1836 that form atom-like entities similar to the atoms of our world.

The point generalizes: Sometimes the apparently arbitrary parameters can be explained by the necessary features in the essences of things, but in those cases it will often be the case that it is contigent that these essences, rather than other similar ones, are exemplified. In those cases, the appearance of arbitrariness yields an appearance of contingency, and the true contingency is nearby.

There is, however, a further worry here. Consider the apparent arbitrariness of the fact that the ratio of the circumference of a circle to its diameter in decimal notation has 1 and 4 as its second and third digits, respectively. Yet this fact can be wholly mathematically

[4] In ??ref, I have argued for a Principle of Sufficient Reason (PSR) that holds that all contingent facts have an explanation. But even if one rejects the PSR, one should hold that explaining relevant contingencies is a good feature of a theory, one that provides evidence for the theory.

explained by necessary mathematical truths such as that $\pi = 4 - \frac{4}{3} + \frac{4}{5} - \frac{4}{7} + \cdots$.[5] Thus the appearance of arbitrariness of a parameter is merely *defeasible* evidence of contigency in the parameter or even nearby.

We thus have to be cautious: moving from apparent arbitrariness to contingency, whether of the parameter itself or of something "nearby", is always going to be a defeasible and non-deductive move. This is why there is a value in Mersenne's giving as many examples as he does, since non-deductive arguments tend to stack up. But in any case, a number of Mersenne's particular examples, such as the astronomical distance examples, are ones where it would be difficult to believe in a necesity-based explanation without any contingency involved.

In the rest of the book we will find that if we turn our attention away from science and towards philosophy, we will find a myriad of cases like Mersenne's where there are seemingly arbitrary parameters. But these will be cases where a randomness explanation is implausible, bruteness is not satisfactory and the appearance of contingency is undefeated. However, unlike in Mersenne's cases, I won't be arguing—at least not in the first instance—that theism provides the solution. Rather, the solution will be Aristotelian metaphysics of form.

[5]This point is very similar to an argument Hume makes in Part IX of his Dialogues??ref.

CHAPTER II

Mersenne questions in ethics

1. Motivating examples

1.1. The rule of preferential treatment. Let us begin with a more detailed discussion of an example from Thomas Aquinas's discussion of the order of charity. Aquinas thinks, along with common sense, that those who are closer to us have a greater moral call on us. Thus, if it is a question of bestowing the same good on one of two people, where one is more closely related to us, we should benefit the closer one. But Aquinas writes: "The case may occur, however, that one ought rather to invite strangers [to eat with us], on account of their greater want."??ref And then he raises the question of what one should do "if of two, one be more closely connected, and the other in greater want."??ref

We might hope that here Aquinas would give us some clever rule for weighing connection against need. But instead he writes very sensibly: "it is not possible to decide, by any general rule, which of them we ought to help rather than the other, since there are various degrees of want as well as of connection".??ref It is tempting at this point to throw up one's hands and simply say that in these in-between cases there is no fact of the matter as to what should be done, or both options are permissible, or else relativism applies to the case. But that would not do justice to the way we agonize when we find ourselves in such a difficult situation, trying to discover the truth of the matter. (It is interesting to note that the most common real-life moral dilemmas tend to be

like these kinds of cases, rather than highly controversial questions about trolleys, strategic bombing or bioethics much discussed by philosophers.) And indeed Aquinas maintains a realist attitude to the question while simply offering this advice for how to figure out the answer in a particular case: "the matter requires the judgment of a prudent man.??https://www.newadvent.org/summa/3031.htm#article2

We can think of this as the problem of specifying a function $f(r, a, s, b)$ of four variables, two of them, r and s, being degrees of relation and the other two, a and b, being degrees of benefit, where the function takes one of three values corresponding to whether it is obligatory, permissible but not obligatory or impermissible to bestow a benefit of degree a on a person with relation of degree r to the agent in place of bestowing a benefit of degree b on someone related to degree s.

In fact, the problem of a rule of preferential treatment is much more complicated than the above indicates. First, the *kinds* of benefit and relation also matter: "we ought in preference to bestow on each one such benefits as pertain to the matter in which, speaking simply, he is most closely connected with us."??ref So the function will depend not merely on quantitative features but qualitative ones. Second, although Aquinas does not mention it here, the evaluation will no doubt depend on various features of the circumstances. And, third, in practice instead of choosing between two certain benefits, we are choosing between two probability distributions over the space of possible benefits.

Now, as Aquinas admits, we do not know what the moral evaluation function for choices between benefits to different people is. But abstractly speaking there is some such function, even if we do not know what it is, just as there is a function that assigns to each person alive now the number of hairs they now have, even though we cannot specify any of the values of the function. And we have good reason to expect the moral evaluation function to be very complicated. Indeed, probably the only serious proposal for a relatively simple function f here is the utilitarian suggestion that $f(r, a, s, b)$ yields obligation when $a > b$, mere permission when $a = b$ and prohibition when $a < b$. But this utilitarian

suggestion betrays the intuition that the degrees of relation r and s, much less the kinds of benefit and relation, are relevant to the moral evaluation.???refs

Indeed, the function is apt to look arbitrary. Fix the degrees of relationship to be one's parent and a total stranger, and fix a specific and certain financial benefit of $1000 to one's parent, and fix the circumstances. Then as we vary the financial benefit to the stranger from zero to infinity, we will presumably initially have a requirement of benefiting the parent (it would be wrong to give $1 to a stranger instead of $1000 to a parent in ordinary circumstances), then a permission either way, and then a requirement to benefit the second party. There will be boundaries between these regions of logical space, and these boundaries will look as arbitrary and contingent as the boundaries between different tax brackets. Like the tax brackets, some proposals for boundaries will be *clearly* unreasonable, but there will be many proposals that appear reasonable. And whatever the actual boundaries will look arbitrary.

Of course, seemingly arbitrary numbers can come out of an elegant and simple rule: it seems arbitrary that the fifth and sixth digits of π are 5 and 9 respectively, but there is an elegant mathematical explanation. But apart from the utilitarian proposal, we do not have any at all plausible simple proposal for f.

These seemingly arbitrary boundaries in the order of charity raise call out for an explanation at least as much as the exact distance between the earth and the moon does. Just as it seems implausible that the distance between the earth and the moon *must* be exactly what it is, it seems implausible to think that the boundaries must be exactly where they are—unless the utilitarian is right about f being very simple.

In fact, the ethics case calls out for an explanation even more than Mersenne's scientific examples did. For we might be able to swallow the earth-moon distance being a contingent and brute unexplained fact. But a brute fact seems unfitting for a moral rule. A claim that it just so happened, with no explanation at all, that you should ϕ undercuts the moral force of the alleged moral obligation. We expect anything seemingly arbitrary in our moral norms to have an explanatory ground.

To further argue for this point, consider a version of Divine Command Theory on which obligations are divine commands, and God rolled indeterministic dice to decide which actions to command, and by chance God's commands coincided with our common-sense morality, though they could just as well as well have commanded cruelty and dishonesty. A Divine Command Theory on which it is mere chance that cruelty is forbidden rather than commanded provides an unacceptable answer to the Euthyphro problem.?? Intuitively, a set of injunctions that is as arbitrary as that cannot constitute morality. But this point generalizes beyond divine command theory. Suppose that that we have some preferential treatment rules that are brute and contingent, and could just as well have enjoined on us the anti-utilitarian rule that we should always prefer the lesser benefit. Then whatever these rules are, they do not constitute morality, but at best happen to agree with morality in content.

Thus, even if there is some bruteness in the rules of preferential treatment, the rules in our world must be generated in a way that makes rules such as the anti-utilitarian rules not be among the possible outcomes. But this makes it very unlikely that the rules would be brute. For what force would limit the brute rules to avoid unacceptable options? Such a view of limited bruteness would be akin to a view on which banana peels can come into existence *ex nihilo*, but not where we might trip over them.

It is important to remember that the Mersenne question here is a metaphysical question: What explanatory grounds are there for why this rule, rather than some competitor, holds? The epistemic question may well have a virtue-theoretic answer like Aquinas's: if we acquire the requisite virtues, we will be able to judge particular cases fairly reliably, and until then our best bet is to ask the advice of virtuous others.

But before I continue the discussion of the possible explanation for the above ethical Mersenne question, let me follow Mersenne's lead and multiply the examples, in order to defend against potential answers that only work in some cases, and to make clear how widespread the problem is.

1.2. Risk and uncertainty. Some people—perhaps you—would accept a 92% chance of winning a thousand dollars at the cost of an 8% chance of losing ten thousand. I wouldn't. I say that both I and they are reasonable. On the other hand, someone who (in ordinary circumstances) rejects a 99.9999% chance of winning a thousand dollars at the cost of a 0.0001% chance of losing ten thousand and someone someone who accepts a 10% chance of winning a thousand dollars at the cost of a 90% chance of losing ten thousand are unreasonable. It is well known that attitudes to risk vary between people, and while there are unreasonable attitudes, it is very plausible that there is a broad range of reasonable attitudes.??refs So, as we vary the probabilities of wins and losses, we move between cases where accepting the risk is unreasonable, to cases where both accepting and rejecting are reasonable, to cases where rejecting is unreasonable.

This, once again, raises the Mersenne problem of why the transitions between the various evaluative categories lie where they do. And of course things are more complicated than described above. The rational evaluation function will depend not just on the probabilities involves but also on the values of the potential gains and losses.

While in the previous case, utilitarianism provided a neat but implausible solution, so too in this case, expected utility maximization provides a neat but implausible solution. On expected utility maximization, you are rationally required to accept a chance p of a good of degree α despite a chance q of a bad of degree β against a status quo of value zero just in case the expected utility $p\alpha + q\beta$ is strictly positive; when it is zero, you are permitted but not required; and when it is negative, you are not permitted. One problem with this solution is it requires all goods to be neatly quantifiable (cf. the next example for difficulties related to that). But the more serious problem is that it requires an implausibly negatively judgmental attitude towards ordinary people's attitudes to risk.

Indeed, here is a plausible trio of theses about risk that are incompatible with expected utility maximization:

(1) There is no upper bound on possible finite utilities.

(2) A decade of the worst tortures the KGB could think of has a finite negative utility.

(3) There is no possible good G of finite utility such that one would be rationally required in accepting a certainty of a decade of the worst tortures the KGB could thing of one for a one in billion chance of G.

For as long as $(1/1000000000)\alpha + \beta > 0$, where α is the value of G and β is the (highly negative) value of the tortures, one would rationally required to accept the deal on expected utility maximization, and by (1) and (2) there exists a possible G that makes $(1/1000000000)\alpha + \beta$ strictly positive. Hence, we should reject expected utility maximization, and and absent expected utility maximization, it is likely that the rationality evaluation function for risk will be messy and arbitrary-looking.

The most plausible thing for the apologist for expected utility maximization to reject is the no-upper-bound thesis (1). Here is one way an argument for such a rejection might go. First, there is a maximum intensity of goods that our brain can handle. Second, goods become significantly less valuable as they are repeated, decreasing in such a way that the sum of the values of any goods you could have over an arbitrarily long life has an upper bound.??refs

But the repetition thesis is only plausible when boredom and other memory-based phenomena are in play. Suppose you have lived for a very long time. Then you suffer from partial amnesia: you have lost all episodic memory of your past meals and of your past pinpricks. You are offered what you are reliably informed is the most delicious and wholesome dessert every prepared by the best chef on earth, a dessert which you are told you've eaten some large number n times in the past, and you may eat the dessert at the cost of a one in ten chance of a small pinprick. It's clearly worth it, regardless of what n is. So now suppose this happens to you every day of a very long life. The marginal value of each such dessert (i.e., the amount it contributes to total lifelong utility), absent memories of past desserts, must be at least one tenth of the marginal disvalue of the pinprick, at least given expected utility maximization. But the disvalue of the pinpricks clearly does not tend to zero with forgotten repetition. Hence, the value of the desserts does not tend to zero. And hence for any finite utility bound, enough such desserts will exceed the bound.

In addition to Mersenne questions about risk and prudential rationality, there will be Mersenne questions about risk and morality. For instance, what risks we may morally impose on others in exchange for a good to ourselves depends in a complex way on one's relationship to these others, the probability of the risk, the degree to which these others accept the risk, the benefit to self, and so on. When I drive, I risk killing other drivers, their passengers, pedestrians by the side road, and so on. But the probability of these awful outcomes is very small, and typically other people on or by the road have accepted reasonable risks (or have had them accepted by proxies, in the case of children), so these dire but unlikely outcomes typically do not render it impermissible for me to go to the grocery store to pick up ice cream.[1] But when the risk is higher, say because I am tired and sleepy after a long day and hence less likely to be a safe driver, the matter becomes less clear. At some point, as the risk increases, it becomes impermissible to go to the grocery store for ice cream. A particularly thorny set of issues arises in the special case of balancing the risk that the innocent are punished with the risk that the guilty go free. And we have the Mersenne question of why the switchovers happen where they do.

Expected utility utilitarians[2] will have a nice answer to this problem. But utilitarianism, as already noted??ref, has many highly counterintuitive implications.

1.3. Orderings between goods. Under ordinary circumstances, it would not be reasonable to choose to be a mediocre mathematician rather than a superb musician. But suppose one's choice is whether to be a superb musician or a superb mathematician? Here we are dealing with incommensurable goods and either choice is reasonable.

But now let's ask this general question: Is it is reasonable to choose to be a mathematician of quality α rather than a musician of quality β? Again, we have a function that takes a number of variables, including α and β and the circumstances, and tells us whether (a) it reasonable to opt to become a mathematician but not reasonable to opt for music, or

[1] I leave open the question whether concerns about global warming render it impermissible.

[2] As opposed to actual-outcome utilitarians who evaluate actions morally based on the actual utilities that would result from an action.??refs

(b) both are reasonable, or (c) opting for music is reasonable but opting for mathematics is not. And, just as before, it is very plausible that the function is extremely complex.

The problem obviously generalizes to all the many kinds of pairings of incommensurable goods there are. In each case, there will be some function of many variables encoding the correct rational evaluation of the situation/, and we will have the Mersenne question of what grounds the fact that this function, rather than one of the infinitely many others, encodes the correct rational evaluation.

We also have Mersenne questions here that involve qualitative rather than quantitative comparisons. Other things being equal, social pleasures are better than solitary ones. This seems rather arbitrary. What makes it be so?

In the preferential treatment and moral risk examples, utilitarianism offered a nice solution. But the problem of incommensurable goods is also going to be a problem for any plausible utilitarianism. Utilitarianism comes in two varieties, depending on whether the good is pleasure or the good is satisfaction of desire. As Mill famously noted, it is essential to the plausibility of utilitarianism that one be able to make a distinction between lower and higher pleasures, so as to get the common-sense conclusion that it is better to be Socrates unsatisfied than to be a satisfied pig.

But once one makes the distinction between lower and higher pleasures, or lower and higher desires, incommensurability quickly shows up, since different kinds of pleasures and desires do not simply come in a linear ranking. Let's suppose that you get more enjoyment and satisfaction of the desire for truth out of mathematics and more enjoyment and satisfaction of the desire for music out of music, and let us suppose (contrary to typical situations) that your choice of life will not affect anyone else. Then it seems right to say that the mathematical and musical lives are incommensurable even on utilitarianism. But even if they are not incommensurable, but equal or one is better than the other, we still have a Mersenne problem as to what level of quality of mathematical life exceeds, equals or falls below what level of quality of musical life. And in fact it will be more complex than that, in that the quality of a mathematical or musical life is clearly multidimensional.

One might try to get out of this by hoping for some precise definition of the degree of pleasure or the strength of a desire. Perhaps there is a neural correlate of the degrees of pleasure or the strengths of desire that can be quantified in a single number. But such an approach is likely to lead to the swinish utilitarianism that Mill wisely rejects. For presumably the neural correlate can be manipulated directly, and the pig could be given pleasures which, in terms of neural intensity, exceed the highest of Socrates' refined joys, and could be made to have a degree of intensity of desire for its swill far exceeding Socrates' desire for virtue.

Moreover, any neural approach is likely to fall prey to questions of cross-species comparison. While pig and human brains are similar, they are not the same, and states of pleasure and desire are likely to be merely analogical. It is clear that some comparisons between human and porcine goods are possible: a tiny human pleasure is worth less than a great porcine one. As one increases the human pleasure and/or decreases the porcine one, there will come cases where neither of the two is to be preferred, and then eventually cases where the human pleasure is to be preferred over the porcine one. But where exactly the cross-over points are is not something we can just read off the neural correlates. And things get even messier when we compare humans to possible beings that have no brains, such as intelligent robots (if these are possible) or aliens with very different biochemistry.

And even if one could give some such precise formulation, we would still have the Mersenne problem of why *this* formulation corresponds with true value rather than some other.

????????ADD: pluralism about values

1.4. A miscellany of other Mersenne questions. There are many other cases which involve thresholds or transitions that appear to be arbitrary.

On strict deontological views, one shouldn't torture one innocent person to save any number of lives. But of course it would be permissible to gently prick someone with a pin to save even one life. Somewhere between the pinprick and the torture is a transition. What makes the transition be where it is?

On threshold deontological views, it is wrong to torture one innocent to save a small number (say, one or two) of lives, but it is permissible to do so to save a very large number (say, a billion). Again, we have a transition to be explained.[3] And note that even if one is a strict deontologist about torturing the innocent, likely one is a threshold deontologist about some other things. Thus, one may think it's permissible to save an innocent life but not permissible to lie to get a deserved (but on other grounds) salary raise, and hence there needs to be an explanation of the grounds of the transition from permissibility to impermissibility. Or one may think it is permissible to trespass on a neighbor's property to save a cat's life but not to save a grasshopper's. Probably everyone who isn't a full-blown consequentialist is a threshold deontologist about some things.

The Principle of Double Effect allows one to foreseeably cause bad effects that it would be impermissible to cause intentionally, as long as these bad effects are not intended either as ends or means. For instance, it seems permissible to bomb Hitler's headquarters even if one finds out that an innocent prisoner is held captive there. But of course there needs to be a proportionality condition imposed on this: the good achieved, say the end of a war, must be proportionate to the bad, say the death of the prisoner. It would be wrong to demolish an old building while knowing that there is a child playing inside: the good of having a lot to build on is not proportionate to the death of the child. So there will be some function of variables including harms and benefits that specifies when the benefit is proportional to the harm in Double Effect contexts. In fact, there will be other variables, such as one's relationships to those harmed and those benefited.

We need to show *respect* for intelligent beings. This respect includes such things as not killing them when they are innocent and non-aggressive, not eating them (except perhaps in extreme circumstances), not acting as if they were fungible, treating them as ends rather than as mere means, and so on. But what is an intelligent being? First, we have a distinction between an individual and a kind based concept of intelligence: on the former, a being is intelligent to the extent that it currently has certain intellectual powers; on the latter, a

[3] I am grateful to Philip Swenson for this example.

being is intelligent to the extent that it is of a kind that should have certain intellectual powers. But whichever we choose, and plausibly there are principled reasons to choose one rather than the other[4], we still have a Mersenne question as to the degree of intellectual power—whether actual or proper to the kind—that is needed for us to have duties of respect. Intellectual powers, after all, clearly come in degrees, and if at some point respect is called for, we need an explanation of why that point shows up where it does.

The question of what in fact the degree of intellectual powers is needed for respect is one that we actually face with regard to our treatment of higher mammals on earth, and that we currently only face hypothetically with regard to extraterrestrial life. It is an important question. But, as usual, the Mersenne puzzle isn't that of determining what the fact is, but of what makes an answer be an answer, especially in light of the appearance that any threshold will be arbitrary.

The laws of a legitimate government should generally be obeyed. But when a government becomes sufficiently unconcerned about the wellbeing of the people, it becomes illegitimate. Why does this transition happen where it does?

Punishment should not be disproportionate to a crime. But in a legal system without a strict *lex talionis*, the proportionality is not going to follow any simple and elegant rule. Nonetheless, there are obvious restrictions. A month's imprisonment for an ordinary parking infraction is disproportionate in one direction; a ten dollar fine for a murder is disproportionate in the other. What grounds the specific rule of proportionality?

Finally, standards of consent necessary to permit one's being treated a certain way vary widely depending on the treatment. There are multiple dimensions in which we can measure the "strength" of a consent requirement: how well informed the consenting party needs to be, what age or level of intellectual development does the party need to have, what proxies if any can offer consent on the party's behalf, how unpressured the consent needs to be, how clearly formulate the consent needs to be, whether the consent must be

[4]Though they will be highly controversial, since a significant part of the debate about the moral status of the unborn turns on this.

specific to the case or whether prior blanket consent suffices, etc. Under ordinary circumstances, no consent—at most, lack of refusal—is needed for a pat on the shoulder. The permissibility of major surgery, however, has a consent requirement of significant "strength" along many of the above axes. On the other hand, the permissibility of sex has a consent requiremnt of even greater "strength" along some of the above axes—thus, while proxy consent and prior blanket consent can suffice for major surgery, they do not suffice for sex.[5] The mapping between the form of treatment and the multidimensional strength of consent is of great complexity, and has an appearance of significant arbitrariness. What grounds it?

Some readers will disagree with a number of the examples. Double Effect, for instance, is quite controversial. But it seems likely that a number of the remaining examples will still compellingly raise Mersenne problems. And the list above is not exhaustive: the reader should be able to generate more items.

2. Arbitrariness

Whatever the values of the parameters in the ethical Mersenne questions are, these values appear likely to be such that if we knew their exact values, we would find them arbitrary. In physics, some hold out a hope that the fundamental constants in the fundamental laws of nature may be "nice numbers" like 2, π, $\sqrt{2}$ or e. It seems intuitively even less plausible that things would so turn out in ethics.

And even if the parameters turned out to be such "nice numbers", that would itself be a very surprising fact, because while such numbers seem very natural in physics, they seem rather less natural in ethics. Imagine that you should benefit your parent over a sibling just in case the ratio of benefits is no lower than $1 : \sqrt{2}$. That would itself seem

[5]It is tempting to explain this in terms of the fact that surgery—or at least the sort of surgery for which proxy consent suffices—benefits the patient regardless of the patient's consent, while sex is only beneficial when consented to. But this is arguably false. Parents can validly consent to an organ transplant between their children, even if the donor is not expected to benefit on balance (though generally there is a benefit from having one's sibling alive!).

arbitrary. It seems that whatever the numbers turn out to be, they will have an appearance of arbitrariness and of contingency.

3. Continuity

Many of the examples involve thresholds, such as the amount of intelligence needed for respect or the degree to which a government needs to care for the common good to have auhtority. It is plausible to reject the idea that there are discrete thresholds, and instead hold that there are continuous functions, say a function $r(x)$ specifying the degree of respect required to be shown to a being with intelligence of degree x.

But then instead of explaining one threshold, one needs to explain the whole complex shape of the "respect function". On the most naive version of this, intellectual power will be graphed along one axis and respect on another, which will raise Mersenne questions about the slopes of the graph, the positions of the inflection points, and so on. But of course in reality, both intelligence and respect have many dimensions, so what we have is a complex function of many arguments and whose values are multidimensional.

In general, ,oving from thresholds to continuous functions only multiplies the degrees of freedom that call out for explanation.

4. The human nature solution

On our Aristotelian picture, the nature of an organism grounds norms about what the organism's structure and behavior should be. In particular, the nature of the organism will ground many arbitrary-seeming norms, such as those governing the range of appropriate sizes of Indian elephants, the migratory behaviors of monarch butterflies, and the lengths of human femurs. Having the nature makes the organism be the kind of organism it is, and imposes on it the associated norms.

In the case of humans, the behaviors include voluntary ones, and so it is unsurprising that there are norms governing these as well. And just as there are many parameters governing bodily structure and sub-voluntary behavior, there are many parameters governing moral behavior, all grounded in the form.

At the same time, Aristotelian optimism provides us with evidence as to what the parameters approximately are. The actual bodily structures of humans give defeasible evidence as to what normative human bodily structure is and the actual behaviors of humans give defeasible evidence of moral norms. And in both cases, we have ways of identifying healthier or more virtuous paradigms, using the optimistic idea??backref that the various ways of doing well tend to hang together with some degree of unity, and the structure and behavior of such paradigms gives us further evidence as to the norms.

Admittedly, there appears to be a disanalogy between health and virtue. We might use a Mahatma Ghandi or a Mother Teresa to figure out moral norms, but we wouldn't use an Usain Bolt or a Serena Williams to figure out physical norms. One explanation of the difference is that Bolt and Williams have highly-developed traits that are specialized to a forms of life quite different from that of the typical human—namely, the life of a professional athlete—while Ghandi and Teresa's excellences in justice, fortitude and mercy are as important to our life as to theirs.

All this raises the question of why the form includes these norms and not others. Here there is an easy answer available. The form is at least partly defined by the norms it includes. Thus, Mersenne's question about the lion and the ant when reformulated into normative terms, as the question of why the lion's strength *ought to* be greater than the ant's, is easily answered: this follows fron defining features of what make lions be lions and ants be ants.

The appearance of arbitrariness and of contingency in the ethical Mersenne problems is somewhat misleading: it is like the appearance of arbitrariness and contingency in the fact that water is H_2O or that carbon atoms have six protons. Water couldn't have a different chemical structure and carbon atoms couldn't have a different number of protons. But it is

also an important truth here that there could be other substances that could have a different chemical structure or a different number of protons. Similarly, *we* couldn't have other norms of preferential treatment than the ones written on our nature, but there could be—and perhaps in this vast universe are—other intelligent animals with other such norms.

5. Other solutions

We thus have many Mersenne questions pointing to arbitrary-seeming parameters in ethical rules. I will now argue that a broad spectrum of ethical theories and solultions are unlikely to yield good answers to the Mersenne questions or else raise new Mersenne questions of their own.

5.1. Kantianism. Kantianism is an attempt to derive moral rules from the very concept of objective rationality. Famously, this leads to difficulties in accounting for the substantive content of rules. For instance, from the point of view of objective rationality, it is difficult to generate a presumption in favor of causing pleasure and against causing pain. The more tightly connected a moral rule is to the specifics of the human condition and of the circumstances, the more difficult it will be for the Kantian to account for it. But the Mersenne questions above thrive precisely on such detail. Consider, for instance, the improbability of a good Kantian account of how much we should, other things being equal, favor siblings over cousins, or of why proxy consent is sufficient for surgery but insufficient for sex. The "logical distance" between the high level principles, like the categorical imperative to treat others as ends and never as mere means or to act according to universalizable rules, and such specific moral content appears unlikely to be bridgeable. Thus, precisely those cases that we have seen to raise compelling Mersenne problems make Kantianism an implausible ethical theory.

Of course, such appearances can be deceiving. One might well have antecedently thought that the relatively simple axioms of set theory are unlikely to generate the richness of mathematical theorems that we have seen to come from them. So it would be good to go beyond an intuition of "distance".

There are at least four ways to do that. First, proceed by intuitions regarding a specific example. Consider two different moral rules regarding to the relative treatment of siblings and cousins. One rule says that benefits to siblings are to be slightly preferred to benefits to first cousins and the second says that first cousins and siblings are to be treated on par. Neither rule requires us to treat anyone as a mere means or takes away from treating people as ends. Both rules are universalizable. So we are not going to be able to derive one rule rather than the other from Kantianism as originally formulated by Kant.

Second, we can makes use of a heuristic as to the validity of arguments. One heuristic I employ in checking whether a numbered argument given by undergraduate students is valid, i.e., whether its conclusion logically follows from its premises, is to see if the conclusion of the argument contains any substantive terms that do not appear in any of the premises. If it does, it is in practice unlikely that the argument is valid, though of course there are possible exceptions. If the premises are contradictory, then the logical rule of explosion makes every conclusion a valid consequence. And it could also be that the conclusion is disjunctive and the substantive term that did not occur in the premises occurs in one disjunct while another disjunct follows from the premises (though I have yet to see this happen in a student paper). An argument from premises about the nature of rationality as such with a conclusion about specific familial relationships or about specific human activities such as sex or surgery fails the heuristic, and hence is unlikely to be valid. And the cases do not seem to be like the most common exceptions—the premises are not contradictory and the conclusion is not disjunctive.

Third, all or most of the examples that raised Mersenne questions have an appearance of contingency to them, in a way that does not fit with the hypothesis that they derive from necessary principles about the nature of rationality. One way to formulate this contingency is to note that many of the rules are ones that we would not expect to apply to other intelligent species. If we came across an alien species that regarded familial ties as somewhat more or somewhat less important than we think permissible for humans, we

should not judge them immoral. It would not surprise us if other intelligent animals—perhaps ones occupying other niches—were rationally or morally required to take greater or smaller risks than we.[6]

Finally, we have an epistemological argument. While clearly we do not know the exact values of the parameters in the Mersenne questions, we have some approximate knowledge, as already indicated above in a number of the cases. We clearly did not come to this approximate knowledge by logically deriving it from Kantian first principles. Nor did we even do so by means of an intuition that they follow from these principles. For I take it that we do not in fact have an intuition that, say, the preference for siblings over cousins follows from Kantian principles. If anything, we have an intuition that it does not. So, it seems that if these rules in fact follow from Kantian principles, it's just a coincidence that our beliefs about the parameters are correct, a coincidence that makes the beliefs be mere justified true belief rather than knowledge. But the beliefs are knowledge. So, the Kantian explanation does not work.

The epistemological argument has some force, but not that much. First, the argument is related to the highly controverted literature on evolutionary debunking arguments.??refs,add?? Second, a theistic reader has an easy way out of the argument: God knows what values of parameters in fact logically follow from Kantian principles and could either directly instil in us correct beliefs about them or ensure that we evolve in a way that yields such true beliefs.

5.2. Act utilitarianism. The main problem with act utilitarianism is that it generates incorrect moral claims. It says that a healthy patient whose organs can save three others can be killed when doing so doesn't have any other countervailing consequences such as making others more callous. It says that if you and I are loners who make no contribution to society, but I own a dog and you don't have any pets, then you have a duty to sacrifice your life for mine, to save my dog from being ownerless; and if neither of us has a pet,

[6]One thinks, for instance, of the Klingons and Kelpians from the Star Trek universe, respectively.

but you enjoy chocolate a little more than I do while everything else is equal, then I have a duty to sacrifice my life for you, since your life would include slightly more utility.

Moreover, as we saw in ??back, for utilitarianism to be plausible and not swinish requires a hierarchy of goods, and there will be Mersenne questions regarding that.

Finally, even hard-nosed desire-fulfillment or hedonistic utilitarianism will be unlikely to be exempt from Mersenne questions. There are multiple mental state concepts that could be argued to correspond to the words "desire" and "pleasure".

When the psychotherapist tells Jones that she always unconsciously wanted to kill her mother, is that a "desire" in the sense of desire-fulfillment utilitarianism or not? A case can be made either way, and this decision point generates a degree of freedom for the theory, and hence a Mersenne question as to why it is one sort of "desire" or the other that counts as defining the good. In fact, reflection the complexity of human life as seen in literature??ref:ColinAllen? shows that there are likely to be many "desire"-type concepts, differing along multiple dimensions, and hence generating a multiplicity of Mersenne question. And there will be multiple ways of quantifying the strength of a desire.

And as for pleasure and pain, we will again have a broad variety of concepts and a multiplicity of ways of quantifying them. This can perhaps best be seen if we think about the mental life of possible and actual non-human sentients. Does a particular state of an earthworm count as a pleasure? It is unlikely to be exactly like a state of ours. There will likely be many ways of classifying mental states across species, and on some the worm's state will be a pleasure and on others it won't. So we have a degree of freedom in our act utilitarianism as to what we count as pleasure or pain in non-humans. And even within humans there are complex questions. Consider for instance masochism or the subtle morose "satisfaction" of the pessimist who sees everything going downhill. There are likely to be different ways of classifying states as pleasures or pains, and the hedonistic utilitarian will have a Mersenne question as to why one rather than another classification is the one that defines ethics.

5.3. Rule utilitarianism. On rule utilitarianism, instead of requiring that each action optimize total utility, it is required that each action follow rules that are themselves optimized for total utility. Rule utilitarianism's main advantage is held to be that its escape from the counterintuitive consequences of act utilitarianism. The rule not to kill the innocent may well be the optimal rule for us, even if in a lifeboat situation it would maximize utility for the two stronger people to kill and eat the weaker third.

Rule utilitarianism could not only neatly explain the apparently arbitrary specifics of the moral rules, but could also explain the appearance of arbitrariness and contingency in a way that, say, Kantianism is unlikely to. For the optimization procedure that would define the moral rules would be a vast and complex one, taking into account the impact of the actions falling under the rules both in the short and the long run, both on humans and on non-humans. It is unsurprising if a complex optimization procedure produces results that seem arbitrary but are in fact carefully chosen to their end. A computer-optimized airplane wing will have precise angles and bends that cannot really be explained without running through the whole computation.

Moreover, rule utilitarianism is less prone than Kantianism to make our limited but true beliefs about the moral rules be merely coincidental. For we have evolved biologically and mimetically in the service of survival and reproduction, and because of the contingent connections between these goods and other aspects of utility, evolution put pressures on us that directed our moral beliefs in a truthful direction. There are deep and difficult questions whether this is enough to make the connection between our beliefs and the truth be sufficient for knowledge??refs, but there is more hope here than on the Kantian side.

However, famously, rule utilitarianism divides into two varieties, depending on exactly what the rules are optimized for. On ideal rule utilitarianism, the rules are such that everyone's successfully following them would be optimal, even if in fact they are too difficult for us to follow. Ideal rule utilitarianism, however, is widely held to reduce to act utilitarianism, since if everyone were to actually follow the rule of maximizing utility, that

would be optimal with respect to maximizing utility. But act utilitarianism has already been put aside.??backref

Non-ideal rule utilitarianisms, on the other hand, inject a note of realism into the optimization procedures. For instance, what might render a set of rules correct is that if everyone were to *try* to follow them, optimal results would result. This already raises a Mersenne question. For trying is something that comes in degrees, and it is very likely that different rules will be generated when we optimize for the utility resulting from everyone's trying hard to follow them than if if we optimize for the utility resulting from everyone's trying with minimal effort. And there will be a vast number of intermediate cases, so there will be a Mersenne question of what grounds the fact that α, say, is the right degree of effort for defining the optimization procedure that generates the moral rules.

Furthermore, specifying the degree to which the hypothetical agents try to follow the moral rules is not enough to specify the optimization procedure. For instance, one has to specify the level of intelligence of the hypothetical agents, their non-moral interests and the environment, which yields multiple Mersenne questions as to what the requisite levels of these for the hypothetical optimization procedure are.

The only way to avoid such questions is to simply require the counterfactual world to match our world in the respects, but this runs into two problems. First, we would normally expect a world where all agents try to follow the moral rules to have agents that have different non-moral interests, higher levels of intelligence since such a world would have a much more just educational system than ours and hence would nurture children into greater intelligence, and a rather different natural environment. If we try to keep the three factors fixed while having the hypothetical agents try to follow the moral rules, we are likely to get some very unlikely counterfactual results, just as keeping too much of our world fixed in a counterfactual situation results in the odd claim that if Oswald did not kill Kennedy, Kennedy would have been buried alive. Second, we have to say that if our history had gone slightly differently, so that (say) the distribution of intelligence in the general population were slightly different, the optimization procedure would have generated

different rules, and hence different moral rules would have been true. Indeed, on this view we would get the very strange idea that what we morally do can affect morality itself.

Besides this, there are other non-ideal aspects that we should probably introduce. Some of our important moral rules discuss how we should deal with culpable malefactors. But in a world where everyone tries to do the right thing, depending on the strength of trying, there might well be *no* culpable malefactors, or at least very few. And it is unlikely that moral rules optimized for such a very different situation would be likely to be the right ones for us. So we probably need to optimize the rules with respect to a hypothetical situation where not everyone tries to follow them. And that raises Mersenne questions as to how many people in the hypothetical case follow these rules, and what the others do with their lives.

In short, ideal rule utilitarianism is implausible, while developing the non-ideal rule utilitarian project raises multiple Mersenne questions as to the details of what is to be fixed in the hypothetical situation.

5.4. Social contract. Contemporary social contract theories are based on duties grounded in hypothetical agreement between agents in situations of ignorance.??refs Anyone who has been in a long committee meeting knows that actual agreement between agents can result in complex rules with much apparent arbitrariness, and it would be unsurprising if hypothetical agreement were similar. Thus far, social contract fits our data well.

But the hypothetical agreement condition involves multiple parameters such as how smart the hypothetical agreers are (and there are multiple dimensions of intelligence), what exactly are they ignorant of, how many of them are there, what are their attitudes towards risk and uncertainty, etc. We have here an explanation of the Mersenne parameters in terms of other Mersenne parameters, and the problem remains fully entrenched.

5.5. Virtue ethics. Aquinas himself invoked the virtuous agent as providing at least the epistemic path to an answer to the preferential treatment question. We could also take

virtue ethics to provide an answer to the Mersenne question: What makes these parameters, rather than others, hold is that the virtuous agent's patterns of behavior are thus and so parameterized.

But this of course simply shifts the problem to that of why the virtuous agent's patterns of behavior are parameterized as they are. The best answer to that question appears to be the one given in the Aristotelian tradition which grounds this in the agent's nature.

5.6. Divine command. On divine command ethics, the right is what is commanded by God. Divine command ethics, like social contract and rule utilitarianism, carries with it significant hope for explaining the apparent arbitrariness in ethical parameters. We would not be surprised if the laws coming from an infinitely intelligent and good legislator had significant complexity that to us would look like arbitrariness.

It may initially seem the divine command ethics runs into the same problem of pushing the Mersenne questions back to the question of why God legislated these parameters and not others. But notice that the Mersenne problems I have been discussing are *grounding* questions. Even if God's legislation were completely arbitrary in a way that ultimately violated the Principle of Sufficient Reason, on divine command ethics we would have a *ground* for the parameters in preferential treatment and other ethical rules being what they are. To say that we should prefer siblings over first cousins in a ratio of $1.7 : 1$ because God commanded so is to give a ground for the obligation, even if that ground itself needs an explanation. Compare the moral prohibition on adding cyanide to friends' drinks. There would be something absurd if that prohibition were ungrounded. But it has a ground, or at least a partial ground: cyanide is fatal to humans. Imagine now that there was in fact no possible explanation of why cyanide is fatal to humans. Nonetheless, the grounding problem for the moral prohibition would have been solved by citing the danger of cyanide.

In this way, our ethical grounding Mersenne problem is quite different from Mersenne's merely explanatory problem. In Mersenne's case to explain why the distance between the earth and the moon is what it is in terms of other parameters of earlier states of the solar system does not make significant progress. But when we have given a plausible

ground to the moral obligation, we have indeed made progress. Mersenne's original argument depends for its plausibility on a fairly general Principle of Sufficient Reason.??ref-on-PSRr Here we just use a heuristic principle that moral truths with an appearance of arbitrariness need a deeper ground.

Moreover, the divine command theorist has nice answers available to the question of why God chose these rules. For instance, God could be an act consequentialist and could have optimized the rules to produce the best consequences, including perhaps such consequences as the value of following and disvalue of breaking moral rules in addition to first order values and disvalues like pleasure and pain. We would expect a complex optimization to produce results with an appearance of arbitrariness. A sailboat hull computer-optimized to minimize drag is likely to have many parameters that look arbitrary to those who do not know how it was generated.

At the same time, we still have some serious Mersenne grounding problems. The plausibility of divine command ethics rests in the idea that God is a legitimate authority and legitimate authorities need to be obeyed. This suggests that logically prior to divine command ethics there is some sort of a proto-ethical general rule about obedience to legitimate authority. That rule itself will have to have parameters specifying which authorities are legitimate and what the scope of their authority is. And we will have the Mersenne problem of grounding these parameters.

Moreover, even if we do not have such a general rule about all authority, but a specific rule about divine authority, this will still raise some Mersenne problems. For, as Aquinas noted??ref, legislation only has a claim on our obedience when it is appropriately promulgated. And promulgation is a complex concept involving thresholds and parameters. It is not necessary for promulgation that all those subject to the legislation have heard of it. But it is not enough for the legislators to meet secretly, and write the legislation on a stone buried on public land. Intuitively, we need the legislation to be reasonably accessible to those governed by it, but there are many parameters hidden behind the word "reasonably", and we need grounds for them all.

Nor is it even the case that the promulgation condition on God's commands is met in a really clear way, so that all that would suffice is some proto-rule that has a really strict and non-arbitrary promulgation condition like that everyone governed knows of the rules. For any such strict condition is likely to have in fact been violated by God's commands, since there is no agreement on what God's commands are—or even on there being a God.

What is worse, when we focus on the Mersenne cases in ethics, it unclear that divine commands instituting the parameters would even satisfy a fairly modest promulgation that requires those who try really hard to be able to find what the legislation is when it is relevant to life. There surely are cases where we have tried really hard to figure out what is the right thing to do and we didn't succeed. Perhaps it could be argued that we didn't try "hard enough", but now we are the true Scotsman territory.??more?

6. Other attempts at escape

6.1. Particularism. One might try to escape the Mersenne questions by opting for particularism. On particularism, while there may be general rules like "Other things being equal, don't torture people", the application of these general rules to specific situations is not rule-governed. Hence, there won't be a rule specifying when one, say, favors a sibling over a cousin. Instead, there are particular facts about what to do in particular situations.

However, particularism only multiplies the Mersenne questions. For whereas on rule-based systems we had Mersenne questions about why the parameters in the rules had the values they do, now we will have Mersenne questions about why in particular actual circumstances C_1 we should act one way while in slightly different particular actual circumstances C_2 we should act a different way.

Furthermore, plausibly, there will still abstractly speaking be a function that assigns to each circumstances a hypothetical determination of how one would be obligated to act in that circumstance. There may, of course, be no formula specifying the function, but that does not affect the Mersenne question of why this function rather than another, perhaps similar one, is correct.

6.2. Brute necessity. Perhaps we could say that it is a brute, unexplained but necessary truth that the answers to the ethical Mersenne questions are as they are. The boundaries lie where they do, but there is no special ontology behind them: it's just a necessary truth that we should prefer parents to cousins, that an armed up-rising up against a regime responsible for Nazi-style atrocities is permissible while only non-violent protest against the faults of modern-day Canada is permitted, and so on.

Of course, brute necessities should never be a first resort in theorizing, but sometimes they might be acceptable as a final resort. Consider Mersenne-type questions one could ask about set theory. If the Zermelo-Fraenkel with Choice (ZFC) Axioms for set theory are consistent, then for every natural number n they are compatible with the hypothesis CH_n that there are exactly n cardinalities strictly between the cardinality of the natural numbers and the cardinality of the real numbers (the hypothesis CH_0 is the famous Continuum Hypothesis).??ref:check Suppose it turns out that in fact CH_{15} is true. We would have an excellent Mersenne question as to why it is CH_{15} that is true, but the mind boggles as to what could be a satisfactory answer to that question, much as it does in the ethical questions. Perhaps the truth of CH_{15} could be a brute fact, albeit a necessary one since it seems implausible that mathematical truths be contingent (though see ??Pruss for an Aristotelian metaphysical story on which they might be).

Some brute necessities can perhaps be admitted in ethics. For instance, if CH_{15} is necessarily true, then it is necessarily impermissible for us to punish someone for falsely informing us that CH_{15} is true. This impermissibility would derive from the impossibility of CH_{15} being false (and hence the impossibility of falsely informing someone of that it's true) and the impermissibility of punishing people for actions that they did not do. (It is possible, of course, to insincerely inform someone of a necessary truth. But that's a different wrong action, even if equally bad.)

But truly ethical brute necessities are deeply implausible. Here is one way to see this. Suppose there is a sequence s of one or more English sentences expressing your favorite set of fundamental and necessarily ture ethical norms. For instance s might be the single

injunctions "Love your neighbor as yourself" or "Maximize total pleasure minus pain of all sentients", or it might be a longer list. Encode s into a sequence of decimal numbers in some natural way, for instance by encoding each symbol in s into a three decimal digit ASCII number. It is widely believed—though it has not been proved—that π is a normal number, so every possible sequence of digits occurs in it. If so, then the decimal encoding of s occurs somewhere inside π—and even if not, it may well still do so. Suppose that the decimal encoding of s occurs in π as the nth through $(n+m)$th digits. Now consider this metaethical theory: (**??**)To do the right thing is to follow the English injunctions in three decimal-digit ASCII encoding between the nth and $(n+m)$th digits of π. Call this π-metaethics. On the hypothesis that the fundamental ethical injunctions are necessary and can be expressed in English, some version of π-ethics has the correct normative content. But, nonetheless, no version of π-metaethics has any plausibility. For there is no plausible normative connection between an injunction being found inside π and its being binding on us.

Admittedly, if we in fact found a sequence of English injunctions near the beginning of π (say, starting with the tenth digit), we would have some reason to follow them. But the reason would be something like this: The best explanation for why these injunctions are found in π is found in a being or beings that in some way incomprehensible to us can control mathematical truths or, more plausibly, the evolution of our linguistic systems, and there is good pragmatic reason to follow the commands of such beings. Perhaps they have our good in mind, perhaps they will get mad if we don't follow their commands, or perhaps they are trying to inform us of the true ethics. But nonetheless π-metaethics would be false. The reason these injunctions would apply to us wouldn't be that they are found in π, but something else, such as that a being with practical authority commanded them to us or a being with epistemic authority informed us of them.

In other words, a metaethics where the ethical claims are grounded in something intuitively of no relevant to our moral activity, such as the content of the digits of π, is not plausible. To be a candidate for a grounds of ethical claims, a thing needs to be ethically

compelling. For a more controversial illustration of this point, consider that no collection of the traditional attributes of God (omnibenevolence, creation, omniscience, omnipotence, etc.) is such as to make it plausible that the commands of a being with those attributes are what ethics is (??ref:MacIntyre??), and this is a strong reason to doubt divine command metaethics.

But now take some attempt at founding an arbitrary-seeming ethical principle on a non-compelling ground, say the digits of π, and remove the ground altogether. Removal of the ground surely does not make the story any better. Someone who said that what explained why we should favor siblings over cousins by a margin of twenty percent by saying that it is thus written starting with the nth digit of π would be ethically ridiculous (though if n is small, finding the injunction might be some evidence for its correctness). But suppose we drop the spurious π-based ground: surely the ungrounded ethical claim is no better off than the spuriously grounded one.

There may be ethical truths that are not themselves grounded. But these truths should be compelling ethically—perhaps the Golden Rule is like that—and not have an appearance of arbitrariness. And there may be arbitrary-seeming truths in ethics, but they are not fundamentally ethical.

6.3. A two-step vagueness strategy. It is very tempting to dismiss the Mersenne questions above with a two-step strategy. In each case, we first give non-arbitrary grounds for an approximate and vague determination of the parameters involved. Thus, while it is implausible to think that, say, social contract theory will generate a precise answer to the preferential treatment question, it is reasonable to think it will generate claims like: "Benefits to siblings are to be *somewhat* preferred to benefits to cousins." And, then, we simply note that the Mersenne question as to the grounds of the exact dividing line has the false presupposition that there is an exact dividing line—instead, we have insuperable vagueness. This is probably the best alternative to the form-based approach I am defending.

A first concern with the two-step strategy is to worry whether other ethical theories can actually generate sufficient non-arbitrary grounds that have the degree of precision

that we think really is there. This concern has two variants. One involves cases where we know what the facts generating the Mersenne questions are. Kantianism, for instance, is unlikely to generate even a vague morally-relevant distinction favoring siblings over cousins, and yet we know there is such a distinction. The problem of ranking types of goods generates difficult Mersenne questions as to what grounds comparisons that we know are there, such as that fundamental philosophical truths are more valuable than the pleasures of chocolate. The second variant of the concern involves cases where we agonize over what to do. Our agonizing is a sign of our intuition that there is an answer to a moral problem, albeit one we cannot discern. While we may not be seeking for absolute precision, and may be willing to accept some level of vagueness, in a number of cases we seek for more precision than the various alternatives to the form-based theory can ground.

A more radical problem with the two-step strategy is a general problem for non-epistemic views of vagueness: the violation of logical principles.??ref:Sorensen Fix circumstances C and observe that:

(A_0) Giving \$1000 to a stranger is better than giving \$0 to one's parent.

Now for each positive integer n, consider the statement:

(A_n) Giving \$1000 to a stranger is not better than giving \$$n$ to one's parent, or giving \$1000 to a stranger is better than giving \$$(n+1)$ to one's parent.

From A_0 and A_1, one concludes by the disjunctive syllogism that giving \$1000 to a stranger is better than giving \$1 to one's parent. From this and A_2, by the disjunctive syllogism one concludes that this is true even if what one gives one's parent is \$2. Continuing onward, once we get to A_{2000}, we conclude that it's better to give \$1000 to a stranger than \$2000 to one's parent, which is false. The argument is valid: the conclusion follows from the premises by repeated use of the disjunctive syllogism. When we have a valid argument for a false conclusion, one or more of the premises has to be true. Since A_0 is true, it must be that A_n is false for some $n > 0$. Thus, by De Morgan, there is an n such that:

(4) Giving \$1000 to a stranger is better than giving \$$n$ to one's parent and giving \$1000 to a stranger is not better than giving \$$(n+1)$ to one's parent.

In other words, classical logic guarantees that there is a transition point between where it's better to give to the stranger and where it's not better to give to the stranger. ??add-stuff-on-intuitionism-and-check-literature

There are two standard ways out of such anti-vagueness arguments. The first is a logic with more than two truth values. Intuitively, the statement

(B_n) Giving \$1000 to a stranger is better than giving \$$n$ to one's parent

is true for $n = 0$ (note that B_0 is just A_0), but becomes less and less true as n increases. But, surely, B_1 and B_2 are simply true, too. On the other hand, B_{999} and B_{1000} are simply false. So it does not seem to be the case that we always have *strict* decrease of truth value with increasing n. And hence whereas in the classical logic reading we had one transition to be explained, from true to false, now we have at least two: from truth to truth values intermediate between true and false, and from intermediate truth values to falsity. And the transitions appear to be just as arbitrary as before. Thus we have doubled the number of Mersenne questions. And if we say that the second-order questions are also taken into account with multivalent logic—say, it's being the case for some n that B_n is neither true nor false that—then the multiplication of questions increases even more.

Perhaps, though, one can dig in one's heels and insist on strict decrease of truth value. But the precise assignment of intermediate truth values—say, B_{505} getting a truth value of $T_{0.51}$—also calls for an explanation. Thus it seems we have a vast multiplication of Mersenne questions. But there is a response to this argument: ??refs argues that the exact truth values are a mere feature of the logical model and all that has reality is their ordering. And the ordering of the truth values is, perhaps, quite non-arbitrary in that B_m is truer than B_n precisely when $m < n$. But the insistence that the ordinal properties of truth values is what has reality still does not escape the multiplication of Mersenne questions. For consider a different set of ethical questions involving a threshold. For instance, let C_x say that one has a duty to obey the orders of a government that cares to degree x about

the common good, for some method of x of quantifying care about the common good, where, say, $x = -1$ corresponds to the Nazi German state and $x = 1$ corresponds to modern Finland. Then C_{-1} is pretty false C_1 is pretty true. But even if all we insist on is the ordering of truth values, then we will still have a vast, perhaps infinite, number of Mersenne questions like:

(5) At what value n does B_n become less true than $C_{0.24}$?

For clearly B_0 is truer than $C_{0.24}$ while B_{2000} is falser.

?? higher levels of multivalued logic

The most common response to vagueness these days is supervaluation. A the terms of a sentence can have multiple precisifications, with a different truth value corresponding to a different choice of precisifications. "Bob is bald" may be true if we precisify "bald" as having less than half a cubic centimeter of scalp hair and false if we precisify it as having fewer than a meter of hair. Then we have vagueness. When, on the other hand, a sentence is true (respectively, false) under all precisifications, we say it is super-true (super-false).

In the ethical examples, such as whether it is better to give $200 to a stranger or $100 to a parent in circumstances C, presumably the supervaluationist escape from Mersenne questions will be that no matter how far we precisify C, the statement will be vague due a vagueness in ethical terms such as "better" or "right" or "wrong" which have multiple precisifications yielding different truth values for the ethical claim. For instance, it may be better$_{17}$ to give the double amount to the stranger but not better$_{40}$. Indeed, on a view like this, we will have cases (precisely specified by means of the monetary amounts and the circumstances C) where for some precisifications of "better" it will be better to give to the parent and for others it will be better to give to the stranger.??explain-better

Just as in the multivalued logic case, this multiplies Mersenne questions. For where previously it looked like we have a transition from its being true that it's better to favor the stranger to its being not true, now we have two transitions: from its being super-true that it's better to favor the stranger (say, when the amount of benefit to the stranger is extremely large) to its being vague whether it's better to favor the stranger to its being

super-false that it's better to favor the stranger. And supervaluating at the next level up—say, supervaluating "super-true"—only multiplies the Mersenne questions more.

But there are some additional problems for the supervaluationist response. A standard objection to supervaluationism in general is that it implies that it is super-true that there is a sharp boundary of "bald": for, given any precisification "bald$_i$", there is a sharp boundary for it. In doing this, supervaluationism explicitly forces the denial of its governing intuition that there are no sharp boundaries.

On the ethical side, there is another serious problem. It is truism that we have reason to do what is better. Truisms had better be super-true. This implies two possibilities with regard to the truism. Either for every precisification "better$_i$" we have reason to do what is better$_i$, or else we need to precisify "reason" and "better" in lockstep when we precisify the truism, so that for every i it will be true that we have reason$_i$ to do what is better$_i$. Neither option is satisfactory.

If for every i we have reason to do what is better$_i$, given the existence of infinitely many precisifications here, it seems that the choice whether to favor the stranger and the parent is governed by infinitely many reasons on both sides. This infinite multiplication of reasons is implausible. Moreover, there is no overall winner here—no reason all things considered—for if there were, then we could raise our Mersenne question with regard to the overall winner, and we would be no further ahead. But saying that there is no on-balance reason here denies the intuition that cases near the boundary are hard cases, that it is a difficult question to figure out whether to favor the parent or the stranger, since as soon as one can see that one is in the vague region, one could just conclude that neither action is on balance required by one's reasons.

But if there are infinitely many ways to precisify "reason", none of them privileged, then this undercuts the very idea of our life being governed in a non-arbitrary way by rationality. It seems entirely arbitrary whether we follow reasons$_{17}$ or reasons$_{40}$ in our lives. Many questions of rationality turn into purely verbal questions as to how "reason" is to be precisified. And the same goes for related terms like "morality" and "virtue".

This does not seem to do justice to the non-arbitrariness that is central to our conception of reason, morality and virtue. The point here is similar to the one raised in ??backref regarding π-metaethics: it would be arbitrary to require obedience to the commands that are found starting with the billionth digit of π.

Finally, consider an epistemicist theory of vagueness according to which there is a true semantic theory that assigns to each term the precise meaning it has in the light of the patterns of our use of that term, but that theory is not epistemically accessible to us. Thus, there is a precise fact as to how much hair one can have and yet have "bald" apply to us, a fact grounded in the patterns of our use of the word "bald", but it is a fact that is not accessible to us. Similarly, there is a precise meaning of "right", "better" and similar ethical terms, a fact grounded in the patterns of our linguistic usage. If the transition between bald and non-bald occurs between 98 and 99 hairs, there is nothing mysterious about the fact that someone with 98 hairs is bald and someone with 99 is not, just as there is nothing mysterious about the fact that a backless chair is a stool.

We will evaluate epistemicism in greater detail in ??forward, and even defend a version of it. However, "better" and "bald" are disanalogous. For there is no problem if the question whether the person with 99 hairs is bald turns out to be a merely verbal question, with the true semantic theory simply settling whether "bald", given our usage, means one precise property or some other precise property. But an analogous claim in ethics is implausible. For it suggests an unacceptable arbitrariness on which if our linguistic practices were somewhat different, we would be using the word "better" differently, and there would be nothing less natural about that usage. If so, then our actions' being governed by the better or the right, rather than by some variant property, would be entirely arbitrary. And that is not plausible, for the same reasons that we discussed in the case of supervaluationism.??expand

6.4. Anti-realism. Retreating from realism in ethics to error theory does, of course, remove all the Mersenne problems in ethics. But the cost is high: it is incorrect to say that genocide is wrong. Moreover, since some of the Mersenne problems involve not just

morality but also prudential reasoning, this requires one to deny the correctness of standard prudential reasoning. But perhaps the most serious problem with the error theoretic solution is that we will have parallel Mersenne problems in other normative areas, such as epistemology (??forward) and semantics (??forward), and the cost of error theory there is very high indeed: indeed one will no longer be able to correctly say that one *ought to* accept error theory.

A more moderate solution is to opt for a form of ethical relativism. Relativism, of course, suffers from serious and standard objections.??refs Perhaps the most obvious is that it justifies an ultra-conservative approach: for if what I (in the case of individual relativism) or my society (in the social variant) thinks is guaranteed to be true, then I or society has no reason to take variant views into account, since if you disagree with me or my society, you're guaranteed to be wrong (from my or my society's point of view).

Moreover, relativism is itself prone to Mersenne question. Consider individual relativism first on which a moral claim is true just in case one believes it. The Mersenne question here will be most obvious if one opts for a view on which belief reduces to having a credence above some probabilistic threshold, say 0.95. For then the relativist view comes down to the thesis that a moral claim is true just in case one assigns it a credence of at least 0.95. But that seems arbitrary. Why should one be obligated to do what one has credence of 0.95 in, but not obligated what one has credence of 0.93 in? So we have a threshold problem.

Many, however, resist the reduction of belief to a credential threshold. But if we do not so reduce belief, we should then see belief as just one positive doxastic state among many such as surmising, being inclined to think, believing, being confident that, and being sure that. Moreover, a little reflection shows that such classifications are too coarse grained to do justice to the richness of our mental life. So we have a Mersenne question: Why are more claims made true by my believing them rather than my surmising them or being sure of them?

And thinking that the problem here just involves degrees of confidence is probably neglecting much complexity in the human mind. There is likely a continuum between fully believing and merely acting as if one believes. Why does moral truth show up in the continuum where it does? Or think of the case when the psychotherapist diagnoses one with a subconscious belief. Either such define moral truths or they do not, and whichever it is, we have a Mersenne question as to why. And consciousness itself may come in degree.

Furthermore, a narrow relativism that just makes those moral claims that we actually believe is very implausible. Suppose I believe that it is wrong to eat animals, and I know that cows are animals, but I do not actually draw the conclusion that it is wrong to eat cows. On such a narrow relativism, it would be wrong for me to eat animals but it would not be wrong for me to eat cows, even though I know them to be animals. This is incredible. So we want to extend moral truth at least to things that clearly follow from my moral beliefs. But probably we do not want to extend it to things that follow in ways that are far beyond our ability to know. For, first, if do extend it thus far, then a Kantian might end up counting as a relativist, since the Kantian may think that moral truths are necessary truths, and that necessary truths follows follow from everything. And, second, it seems that this loses sight of the internalist motivations of relativism. But if we restrict moral truth to things that follow *sufficiently easily* from our beliefs. And we will have a Mersenne question of grounding where the line of sufficient easiness lie.

If our relativism is of the social sort, we will have analogues to the above Mersenne questions raised by belief and consequence. And we will have more Mersenne questions There is a complex and difficult literature on how to attribute doxastic states to a community. A reasonable reading of that literature is that there is a multiplicity of concepts that can be expressed with a phrase like "The committee believes that p." For instance, belief by the vast majority of the committee members is enough on the more reductive concepts, while on more procedural versions of the concepts the committee's belief requires some sort of a joint procedure, such as a vote. There will be many answers here, corresponding to a broad spectrum of takes on what a community's beliefs is. And a social relativist will

then have a Mersenne question as to why moral truth is defined by the particular take in question.

The second set of Mersenne question arises from the question of identifying what counts as one's community. I am a citizen of two countries and a permanent resident of a third. Are the moral beliefs of all of these communities—no doubt, mutually contradictory in various ways—true for me, or only of one? Do moral beliefs come to be true as applied to me because I am legally a member of the community, or because I identify with it emotionally, or because I would like to identify with it emotionally? Or is it, perhaps, that every community's beliefs are true for the community, but there is no such thing as being true for the members of the community? (That would nicely solve the problem of contradictions between the various communities I am a part of.) How large does a community have to be to define moral truths? Is a chess club a commmunity that defines moral truths? What if the chess club goes down to one member? Are a pair of friends a community? It is clear that there are many degrees of freedom in a social relativistic theory, and we would have a Mersenne question corresponding to each of them.

CHAPTER III

Ethical and metaethical advantages

1. Metaethics

Metaethics is an account of why the most fundamental ethical truths are true. If we were to make a wish-list for metaethics, it would arguably include the following desiderata for what should follow from the theory:

(6) Ethical truths are objective

(7) Ethical truths are knowable

(8) The explanation of fundamental ethical truths makes them morally compelling to us

(9) The normative implications are plausible.

Recall our π-metaethics on which what made ethical claims true is that they were encoded at some specific position in the digits of π. This gave us objectivity and knowability (at least given the specific position and encoding system).

An individual relativism that says that the right is what agrees with one's belief as to what is right, on the other hand, gives us knowability, and insofar as we find morally compelling the idea that we should obey our conscience it gives us some compellingness, but it lacks objectivity. Furthermore, its normative implications as to what I ought to do are very plausible to me, since obviously I find my own moral views plausible, but the theory's implications for what Hitler should do—namely, that precisely those actions that he believes are right are the ones he ought to do—are implausible to me (and you) in light of the odiousness of his beliefs.

Utilitarianism, on the other hand, considered as a metaethical theory about the nature of the right, yields objectivity, knowability and moral compellingness (the idea that what

we should do is maximize the good is among the *prima facie* most plausible of moral ideas), but it yields a lot of very implausible normative consequences.

A Natural Law metaethics on which for an action to be right is for it to be a proper exercise of the will according to our nature yields a limited objectivity: it makes the right be relative to our kind. But as we saw in Chapter II, this degree of relativity is highly plausible: it is plausible that ethical requirements do vary between different kinds of intelligent beings.

The Natural Law metaethics yields knowability when we accept the Aristotelian harmony theses that things generally function correctly and that the various norms for a thing tend not to conflict. For instance, given such a thesis, the norms for our emotions—including emotions such as moral repugnance or moral admiration or the feeling of obligation—are likely to cohere with our norms for our actions, and by and large our emotions and actions are apt to be correct. This enables us to evaluate normative ethical theories according to the constraint of whether their requirements fit sufficiently with our emotions and require actions that are not too distant from those that people actually perform, especially in the case of people whose lives appear to be harmoniously flourishing. We thus have a rational equilibrium epistemology for our ethics.

The basic idea here is that we ought exercise our will correctly. This is so compelling that it smacks of triviality. Nonetheless, the claim is not trivial, since it provides an analysis of the moral ought in terms of the functional correctness of our wills. We find compelling the idea that we should be true to ourselves. But to be true to ourselves is not just, as is popularly supposed, being true to our changing beliefs and values, but it is to be true to that which makes us be the kinds of things we are: our nature.

The metaethics of right action as the proper functioning of the will is *prima facie* compatible with a very broad variety of normative theories. It seems we can imagine a being whose will's proper function is to will maximal total utilityThus, Aristotelian metaethics is compatible with utilitarian normative ethics, but not with metaethical utilitarianism on which the right is *defined* as what maximizes utility., or to will in accordance with God's

commands, or to will what is universalizable, or to will one's flourishing, or even to cause maximal harm to self. Some of these views will, however, be less plausible given other Aristotelian commitments, such as harmony theses. The harmony theses make it unlikely, for instance, that the right thing be maximal self-harm. Indeed, harmony theses ensure that the normative consequences of the ethical theory be, by and large, fairly intuitive. At the same time, there is a real possibility of error, and of correction of that error.

Natural Law metaethics does justice to the idea that the source of our obligations is in us, rather than in some external fact—such as a divine command—whose moral relevance is questionable. We are our own moral legislators, but because our nature is metaphysically not up to us, we do not have a choice as to what we legislate and we can be wrong about what we have in fact legislated. Natural Law metaethics will thus accept with modifications both the relativist's and the Kantian's insistence on autonomy, but without the ultra-conservative consequences of relativism on which we are always guaranteed to be right and hence never have reason to change our views, and while avoiding the merely formal character of Kantianism which makes it unlikely to yield sufficient normative consequences to guide our lives.

Other metaethical theories may satisfy the four desiderata as well.

2. Flourishing

A substance flourishes to the extent that it functions in accordance with its norms. Acting morally rightly is a case of functioning in accordance with the norms for the functioning of the will. Thus, acting rightly morally is an aspect of flourishing for those substances that have a will. At the same time, unless we should deal with a substance that consists of nothing but will, there will be other aspects of flourishing. Because of this, conflict between moral rightness and self-interest is in principle possible.

Admittedly, Aristotelian harmony tends to limit such conflict. Right action tends to promote other aspects of a substance's well-being. But nonetheless just as jogging sometimes promotes cardiac wellbeing at the expense of joint wellbeing, so too right action promotes volitional or moral wellbeing at the expense of life and other goods.

Starting with Socrates, Western ethical reflection has often insisted on moral wellbeing being the most important aspect of a human's well-being. This may seem to be necessary for preserving the idea that one should do the right thing even when this costs one heavily, by allowing one to insist that the cost of doing wrong is always greater than the benefits. But the thesis that moral wellbeing trumps other forms of wellbeing is neither necessary nor sufficient for preserving the need to act rightly.

It's not sufficient since even if moral wellbeing trumps other forms of wellbeing, there are imaginable situations where doing the right thing will on balance very likely harm one's wellbeing. For instance, suppose I am a bank employee of mediocre morals and the best empirical evidence available to me shows that taking an evening ethics class from Professor Kowalska would be deeply inspiring and turn me into a vastly better person. Unfortunately, the only way I could afford the tuition is to embezzle a thousand dollars from a billionaire's account. This embezzlement is wrong, but I can reasonably expect to be a much better off morally from it.

And it's not necessary that moral wellbeing trump other forms of wellbeing, because if morally right action just is action in accordance with the norms for the will, then it is clear apart from any trumping thesis why morally wrong action is defective: it is defective because it fails to be an instance of the proper functioning of the will. One may be the better off if one does the wrong action, but one will still have acted defectively.

Moreover, it is unlikely to be true that moral wellbeing always trumps other forms of wellbeing. Suppose the best science shows that on average there is on the whole a moral improvement—perhaps in the area of compassion—from suffering severe headaches, but this improvement is tiny. A parent who knew this should still relieve a child's severe headache, and wouldn't be acting contrary to benevolence in relieving it.

Nonetheless, it is plausible that moral wellbeing is typically the most important aspect of our wellbeing and that typically other forms of our wellbeing are appropriately sacrificed to it. This gradation is itself encoded in the human form which specifies what is good for us and the ordering between the goods.

Traditionally, Aristotelian action theory has insisted that we always act for our happiness. This happiness thesis is compatible with a metaethics on which right action is the proper functioning of the will, but is neither entailed by it nor particularly plausible. While proper functioning is always good for a substance, a substance when functioning properly in some way need not be doing so *in order to* function properly in that respect. When a flower opens up in the right season, its opening up plausibly has as its end the good of reproduction rather than the good of opening up. Similarly, when you make dinner for your child, your right action is good for you, but you are doing it for the sake of your child and not for the sake of the action itself.

3. Supererogation

??cut?

CHAPTER IV

Applications

1. Natural relationships

1.1. Siblings and cousins. An interesting test case for ethical theories is whether they can make good sense of our duties to our siblings and cousins. Duties to friends and spouses plausibly arise from commitments we make. Duties to parents have traditionally been grounded in our obligation of gratitude for our life. Duties to children can typically be grounded in the decision to perform actions that have a non-negligible probability of producing a person dependent on us. Duties to strangers might be grounded in our shared rationality. But we owe more to our siblings and cousins than we do to strangers, even though typically we had no say in whether we were to have siblings and cousins, and even when we have no favors to return.

On utilitarianism, our duties to siblings and cousins come mainly from the contingent fact that we tend to be better positioned to do good to them, say because we know their needs better, are likely to be physically closer, and help from us is likely to be more welcome. But if such contingencies are all that is involved, then we also have to accept an error theory about our intuitions when they go beyond these contingencies. If a sibling or a stranger is drowning, other things being equal one should try to rescue the sibling, even if the stranger is slightly easier to pull out, or is likely to have a slightly better future life. If one finds out that a local homeless person is a cousin one has not seen since early childhood, it is more vicious to ignore their needs than to ignore similar need in a random stranger. Murder of a stranger is evil, but fratricide is worse.

In general, utilitarianism, contractarianism and Kantianism focus on the agent's rationality, taking the details of the agent's humanity to provide no direct normative input

into ethical decisions. The fact that most humans hate eating mud gives one reason not to feed mud to them, and the fact that we are unable to instantly teleport ensures we do not have the same obligation to those on other continents as to those nearby. But these are non-normative facts, and the normativity of the conclusions here comes from general normative considerations applicable to all rational beings. There is some *prima facie* plausibility to the idea that the non-normative facts about the relationships between parents and children, together with normative facts applicable to all rational beings, could explain distinctively filial and parental duties. But this is not plausible for the cases of siblings and cousins.

However, if we see ethics as based on the norms written into our *human* nature, given a harmony between the rational and animal aspects of this humanity, will very plausibly allow for distinctive ethical norms tied to particular kinds of natural human relationships, including perhaps in the first instance familial ones. There is no need on our Natural Law ethics to derive the duties to cousins from non-normative facts about cousinhood and norms for all rational beings.

??????????

1.2. Marriage.

2. Double Effect

3. Medical ethics

4. Environmental ethics

5. Relationship to other animals

6. A great chain of being and the definition of life

??move?? Here is an intuition that until fairly recently would have been widely shared: There are deep metaphysical divides between non-living and living things, and between merely living things and persons, and these divides mark a hierarchy of value, a chain of

being. If we could defend such a divide, it would dovetail with the idea that persons are in an important way *sacred*, having rights while other things have mere interests, if that.

I want to offer a highly speculative Aristotelian reconstruction of this intuition. To introduce the reconstruction, start with a puzzle for Aristotelian views. It seems that on such views:

(10) Each thing naturally strives for its own perfections.

(11) The natural activity of a thing is a perfection of it.

But this generates a regress. Let's say that reproduction is an oak tree's perfection. Then by (10), the oak tree naturally strives for reproduction. This natural activity of striving for reproduction, by (11), is then itself a perfection of the oak tree. Therefore, by (10), the oak tree must naturally strive for it: hence the oak tree naturally strives for striving for reproduction. And so on, *ad infinitum*. But surely an oak tree does not pursue infinitely many things. And even after a few level of meta-striving we exhaust plausibility.

I suggest that we can deny (10). Some perfections of a thing are not actually naturally striven for by the thing.[1] The oak tree does strive for reproduction with its reproductive organs. Moreover, it has a second order striving: it strives to strive for reproduction, by growing the reproductive organs with which it strives for reproduction. There may be one or two more meta-levels, but at some level we can say: it just does this, without striving to do it.

Non-living things, on an Aristotelian metaphysics, also have form and also strive for ends. But plausibly they don't strive to strive: they just strive. We thus have a hierarchical division between inorganic things which do not strive to strive and living things which have second order teleological strivings.

The problem of the definition of life is a thorny conceptual problem in biology or its philosophy. Different authors give different lists of features such as homeostasis, growth

[1] An interesting theological example may be the idea in the Thomistic tradition that both the beatific vision and our striving for it are gifts of God's grace, rather than natural for us, even though the beatific vision perfects us.??

and reproduction as part of the definition of life. The multiplicity of features listed makes the concept of life seem arbitrary. Moreover, it is philosophically problematic to tie the the concept of life too tightly to the physical forms of life around us. For it is very plausible that if there are immaterial agents such as deities, spirits or angels, they should also count as alive.[2] After all, those who believe in such beings sometimes hold them to be immortal. But if they were not alive, their immortality would be a trivial claim: a being that is not alive in the first place cannot die. However, these beings are conceptualized as alive, even when they cannot engage in homeostasis, growth or reproduction. And yet while a particular existence claim about the existence of immortal immaterial agents might be false, it does not seem to be fundamentally conceptually confused. Thus, a good account of life should include the kind of life that is attributed to immaterial agents, and none??check of the accounts in the philosophy of biology do that.

Furthermore, it is a merit of a definition that when applied to cases where we do not know how to classify a thing, the definition does not trivially decide the issue, but it points to the question we need to answer if we are to decide the issue. To that end, consider two borderline cases: viruses and sophisticated robots, like Star Trek's Data. In neither case are we confident whether we have life. Viruses are famously a borderline case. And while Data is described as a "synthetic life-form"??ref, and the Star Trek canon clearly favors his being actually alive, the question is not so philosophically clear. Data obviously fails typical biological definitions of life: while he engages in self-maintenance, he doesn't grow or reproduce in the biological sense of the word (though he does make other androids), in a way that does not match typical viewers' intuitions.[3] And whether a virus qualifies as alive varies from definition to definition??ref in a way that makes it sound like the question of

[2]It is worth noting that not everyone who believes in deities, spirits or angels believes them to be immaterial. The ancient Greeks did not think their deities immaterial. And a minority opinion among Christian theologians held angels to be made of "subtle matter".??ref But the argument only needs that some do believe them to be immaterial.

[3]Though, admittedly, there may be some static due to the show confusing the question of consciousness with that of life.??check

viruses being alive is merely verbal. Yet given the strong intuition that there is something of great value about life, even something sacred, the question of what is and is not alive should not be merely verbal.

On the other hand, an account on which what it is to be alive is to have a second order teleological striving—to strive to strive for a perfection—will nicely include any immaterial agents. It will include any entity that prepares itself for future teleological activity, say by growth, and hence will include all the physical forms of life we know about. It will exclude elementary particles. And whether it includes viruses or sophisticated robots is unclear—as it should be. For it is unclear whether viruses and sophisticated robots have form at all. If viruses have form, then it is likely that their activity of attaching to hosts for purposes of future replication is a striving for replicative striving, and hence they are alive. But it is not clear whether they have form. If sophisticated robots have form, they also exhibit meta-striving, and hence are alive. But in both cases we do not know whether there is form, or whether we are dealing with a mere agglomeration of particles. Aristotle himself seems to have thought that artifacts only had form in the analogical sense of a blueprint in the mind of the designer??ref, but he could have been wrong in the case of artifacts like Data. (For more on the epistemic issues here, see Section **??** in Chapter X.)

We thus have two levels in a chain of being: things that strive but don't meta-strive, and things that meta-strive. Now, among the things that meta-strive, we can describe a higher kind of thing: a thing that strives for all of its perfections. The premises of the regress argument with which we started this section apply to such a being. Thus, this is a being that strives for striving for ... for perfection, at any number of levels. While this is implausible for an oak tree or even a dog, we do actually know of one kind of being that does that: humans. Human beings not only conceptualize particular perfections, such as friendship or striving for striving for health, but they conceptual perfection as such, and strive for it as such. If a trustworthy being offered you to increase some perfection or other, and assured you that you would in no way be harmed, it would be rational for you to accept the offer, because perfection as such is one of the things you and I pursue.

At the same time, in a minded being, the infinite chain that results from striving for all one's perfections need not be a chain of separate desires and hence does not require a being that is actually infinite. Rather, all that's needed is for the being to be such that it has or teleologically strives to have the concept of a perfection as such and a desire for perfection as such. This desire then can manifest in a striving to figure out what the perfections are—a striving that is central to the search for happiness (*eudaimonia*) that was so characteristic of Socratic and post-Socratic Greek philosophy—and a striving to be ready to accept whatever one finds. In fact, it might be that for reasons having to do with the nature of infinity *only* a minded being can pursue an infinite number of ends—for any non-minded being that did that would need to have infinitely many distinct causal sources of its activity in a way that might well violate causal finitism, the thesis that it is impossible for an infinite number of causes to work together (for a defense of causal finitism, see ??ref). And among minded beings, perhaps it is definitive of *persons* that they pursue all good.

We thus have a qualitative hierarchy of being between the mere strivers, the mere meta-strivers and the universal strivers. The first division in the hierarchy may well correspond to that between the non-living and the living, and the second might—depending on speculative questions about infinity—align with the division between mere life and personhood. And it is very natural to see qualitative divisions of value here as well.

CHAPTER V

Epistemology

1. Priors

2. Testimony

3. Infinity, self-indication and other limitations of Bayesianism

CHAPTER VI

Mind

1. Naturalistic options

1.1. Multiple realization.

1.2. Functionalism and malfunction.

2. Teleology and representation

3. Teleology and mental causation

4. Soul and body ethics

CHAPTER VII

Semantics

1. A sharp world

CHAPTER VIII

Metaphysics

CHAPTER IX

Laws of nature and causal powers

CHAPTER X

Harmony, Evolution and God

1. Explaining harmony by natures and evolution

[1]

1.1. Number of natures.

1.2. Nomic coordination.

1.3. Aristotelian optimism revisited.

1.4. Fit to DNA.

1.5. Fit to niche.

1.6. Nature zombies.

1.7. Exoethics.

1.8. Epistemology of normativity and form. [Argument: If a guided missile has form, it's alive by the Ch?? account of life. But it's not alive. So it lacks form. Is this a bad argument???]

2. Explaining harmony theistically

3. Explanations of moral norms

3.1. Global aesthetic-like features. [2]

[1] This section owes much to discussion in my mid-sized objects seminar, and especially to Christopher Tomaszewski's suggestions on the explanatory powers of forms.

[2] I am grateful to Nicholas Breiner for drawing my attention, in the context of justice, to this form of explanation of moral features.

3.2. Family.

3.3. Retributive justice.

CHAPTER XI

Eternal Life and Fulfillment

CHAPTER XII

Aristotelian Metaphysical Details

1. Introduction

2. Individual forms

Recall the debate whether forms are individual—numerically different ones for different members of the same kind—or shared by all members of the same kind.

In ??backref, we saw that there is some advantage to an individual form account of ethics: individual forms intuitively do a little more justice to the personal nature of ethical obligation.??[but conjoint twins] ??add But are there any other arguments for taking forms to be individual?

I believe so. An initial attempt might be to argue that then the numerically same entity—the form—is present in multiple places at once. I do not find this argument compelling, however, as I do not think multilocation is absurd.??ref But if you do, that is one argument. Let us consider some others.

2.1. Distant conspecifics. Suppose a shared form theory is true. Now, imagine that in our galaxy there is only one human being, Adam, and imagine that in a galaxy far, far away, God creates a humanoid comes into existence, with no genetic connection to Adam, but with a form that is just like Adam's: this form unifies matter in the same way as Adam's form does, it imposes exactly the same norms on the form's owner as the human form does on Adam, and it causes the same structure and behavior as the human form does for Adam.

At this point we have a dilemma: either the form of this humanoid must be numerically the same as Adam's or not. Suppose it must be numerically the same as ours. Then

somehow simply by creating something in a galaxy far, far away, God causes an entity in *our* galaxy—Adam's form—to become multilocated. This seems counterintuitive.

Suppose that the form does not need to be numerically the same as Adam's. In that case, we have admitted that there can be numerically different forms with the same broadly functional features (including the normative functions). This means that the question of whether you and I have the numerically same form is not settled by noting that the forms have the same functional features. Indeed, now the question whether your and my form is numerically different or the same becomes a metaphysical question that no empirical data is relevant to the settling of. There is nothing absurd about there being such metaphysical questions. But it is some advantage to a theory if raises fewer such questions, having fewer degrees of freedom. And if one does accept a theory where it is possible but not logically necessary that different individual substances have numerically different forms, then one really shouldn't be accepting that in practice you and I share a form. At best one should be agnostic on this question.

2.2. Ethical counting. ...forms are the most important, so why not count by forms rather than individuals, especially in cross-species contextst??