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, Sherif Girgis, Philip Rand, ...??

??Add precise formulation of hypothesis, with various ingredients, and in conclusions add discussion of pieces of hypothesis.

CHAPTER I

Introduction

1. Introductory remarks

I have a human nature or human form that governs my activity, both voluntary and not. 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 great-great-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.

2.3. Species where most organisms fail to reproduce.

2.4. Who are the humans?

2.4.1. *Rational animals.* It is traditional in the Western philosophical tradition to describe the human being as a rational animal. One can take this further and *define* the human being as a rational animal. A number of modern-day Aristotelians??refs do this and hold that if rational dolphins or octopuses evolved, they would also be human in the philosophical sense, having the same nature as we do. While Aristotle certainly held that we were rational, he did not *define* us by our rational animality. And such a move would fit poorly with Aristotle's hierarchical way of defining species that inspired the Porphyrian tree. For if some but not all possible primates were human and some but not all possible

cephalopods were human, then *human* couldn't be a subdivision of vertebrate or of invetebrate. Rather, it would be a species that cuts across multiple genera. This might not bother those of us who do not accept the hierarchical mode of definition[3], but it would be highly problematic for Aristotle.

But there is another serious related problem. The human nature specifies what is normal for human beings. For a rational primate, having four limbs adapted to movement on land is normal; for a rational dolphin, having flippers and a tail would be normal; for a rational octopus, having eight tentacles would be normal. This suggests that there isn't a single human nature that would be shared by rational primates, cetaceans and cephalopods.

However, this argument is much too quick. After all, the variety of normalcy problem seems may well arise even within the biological species *Homo sapiens*, though particular examples are apt to be controversial. The American National Institutes of Health describes lactose intolerance as "an impaired ability to digest lactose"??ref:https://ghr.nlm.nih.gov/condition/lactose-intolerance, which suggests that lactose intolerance is abnormal (we would not talk of an impaired ability to digest cellulose in humans). Yet this alleged impairment is found in the majority of the human adult population worldwide, especially outside of Europe. We could say that speaking of lactose intolerance as an impairment is just a mistake. But there is another possibility: it may be that lactose intolerance is normal for some members of our species and abnormal for others, and the description of it as an impairment is correct when restricted to persons of certain European ancestries. In that case, assuming that all *Homo sapiens* do share the same nature, what makes it abnormal for an adult of Irish ancestry to be lactose intolerant but normal for an adult of Armenian ancestry[4] cannot just be our shared human nature.

Instead, it might be that our shared human nature encodes some conditional like "If you have the complex feature *F*, then you should be able to digest lactose through all

[3]However, see Koons??ref for a fascinating defense of the Porphyrian tree.

[4]There is 4% prevalence of lactose intolerance in Ireland and 98% in Armenia. ??ref:https://www.thelancet.com/journals/langas/article/PIIS2468-1253(17)30154-1/fulltext

your life." This conditional applies to humans worldwide, but only a minority of humans actually exhibit *F*. This feature might, for instance, be genetic coding for life-long lactose digestion, so that those who have this coding ought to be lactose tolerance and those who do not need not be. In that case, even among the Irish lactose intolerance need not always be an impairment: it would only be an impairment amng those who have the genetic coding for lactose tolerance but are nonetheless for some other reason intolerant. Or this feature could specify some other aspect of the genome that correlates with, but does not cause, lactose digestion.

An even more controversial and politically charged example could be sex-linked traits: for some members of our species, having a uterus may be normal, and for others it may be abnormal. Again, our human nature could encode a conditional like "If you have feature *G*, then you should have a uterus." Note that while it is controversial whether there is such a conditional, holding that there is does not by itself decide important questions about gender identity. For instance, someone who thinks that gender identity is determined by genetics could say that the conditional holds with *G* being a genetic feature (say, having two X chromosomes), while someone who thinks that gender identity is determined by personal self-identification could say that the conditional holds with *G* being self-identification as a woman. They could then both use the conditional to argue for opposite answers to the question of whether hysterectomy for sex-reassignment purposes should be covered by health insurance.

Much less controversial are more temporary conditions that affect what is normal. A high temperature is normal for a sick human but abnormal for a healthy one. Yet we surely don't want to say that getting sick is a substantial change.

The problem of the variety what is normal among logically possible biological species of rational animals could be handled analogously. Human nature could specify that if you are a primate you have four limbs, if you are a cetacean you have flippers and if you are an octopus you have eight tentacles. But while this is theoretically possible, it would mean that human nature would have to encode an infinite number of such conditionals. Indeed,

taking this to its logical conclusion, we would have to include conditionals to fit not just rational animals in worlds with laws of nature like ours, but in worlds with a physics radically different different from ours. It seems very implausible to suppose that there is such infinite normative complexity in our nature.

Moreover, the expansive view of humanity results in a very implausible restriction of the space of possible natures. For while Aristotelian harmony may not allow for every combination of normative features—such as an animal that ought to live all its life deep underwater but also ought to breathe atmospheric oxygen—we would expect there to be a broad selection of logically possible natures harmoniously combining different normative features. Thus, just as it is possible to have an animal that is supposed to both echolocate and fly (the bat), and animal that is supposed to both echolocate and be snake-like (perhaps this is unexemplified, but possible), it should be possible to have an animal that is supposed to be both snake-like and rational. But such an animal would not be human, because humans are not supposed to be snake-like (on the expansive view, they can be non-defectively snake-like, while on narrower view, being snake-like would be a defect). So it should be possible to have non-human rational animals.

It should be noted that even if dolphin and octopus persons are not human, it is very plausible that our human nature would require us to treat them with the respect that persons deserve. After all, even on the broad definition of humanity, disembodied beings like deities or angels would not count as human, and yet many people who have believed in such beings have held that we should show them at least the kind of respect we do for fellow humans, for instance by keeping promises made to them.

2.4.2. *The narrower view.* Even within the narrower view of humanity that would exclude dolphin and octopus persons from being human, we still have the difficult question of where exactly the boundaries of humanity are to be drawn. We can make use of our best ethical intuitions to give us good reason to draw them in a way that includes at least all members of Homo sapiens. There was a small study done by philosophers of the motivations of rescuers of Jews from Holocaust, and the one common factor found in the rescuers

was a tendency to identify others as fellow humans. The intuition of these rescuers, as well as of typical anti-racists and anti-sexists in our time, supports taking all members of Homo sapiens to be human.

But there are two further questions. One is about other historical species closely related to us, such as Neanderthals and Denisovans. While we could identify humanity with Homo sapiens, we need not. Probably, we do not know enough about Neanderthal and Denisovan life to see whether it is plausible that one form encompasses the norms for them as well as for us. And fortunately for us this question is of merely theoretical importance.

The second question is whether while we should accept all *adult* members of Homo sapiens as human, we should do so likewise for those biological humans who do not satisfy the philosophers' criteria for developed personhood, such as language or generalized problem-solving skills.??ref:Warren In the case of adults, there is a simple Aristotelian argument for doing so. Adult biological humans who do not fulfill such criteria are abnormal, as is made clear by the fact that if we could treat them medically in a way that leads to the fulfillment of the criteria, we would have good reason to do so. Indeed, if these biological humans were not of the same kind as us, then such medical treatment by transforming them into beings like us would constitute what Aristotelians call a substantial change, a change from one kind of thing to another. But objects do not survive substantial change. Thus, such treatment would be a killing. That is implausible.

The remaining subquestion is for immature members of Homo sapiens, such as zygotes, embryos and fetuses. Here, the questions become more controversial. On a view that excludes such immature members, we would have to say that their ordinary course of life is that they mature and die *in utero*, giving rise to a human in the metaphysical sense. Yet death seems to be a catastrophic event for a substance, and in the growth of these immature organisms into more mature ones there does not appear to be such a catastrophe. This suggests that these members of the our biological species are also human in the metaphysical sense, but much more needs to be said.[5]

[5]??refs

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[6]

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

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

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 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.[7] 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

[7] 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.

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 explained by necessary mathematical truths such as that $\pi = 4 - \frac{4}{3} + \frac{4}{5} - \frac{4}{7} + \cdots$.[8] 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.

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

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.

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.

For a different example, not involving radically large utilities, imagine this Star Trek plot. Captain Kirk visits a planet inhabited by two intelligent alien species, all on par with respect to moral value: there are $1,000,000$ oligons and $2,000,000,000$ pollakons. Unfortunately, an asteroid is heading for the planet. If nothing is done, it will hit the planet in such a way as to wipe out all the pollakons. The only thing Kirk can do is to fire phasers at the asteroid. Spock has calculated that if this is done, the asteroid's track will be redirected in such a way that it will wipe out all the oligons. Kirk asks whether that will help the pollakons? Spock's answer is that probably not: there is a $999/1000$ chance that all the pollakons will still die, but there is a $1/1000$ chance that they will all survive.

It seems very plausible that Kirk should not fire phasers. And it is even more plausible that Kirk is not required to fire phasers. He should not sacrifice the oligons for a small chance of saving the pollakons. But the expected utility of firing phasers is $-1,000,000 + (1/1000)(2,000,000,000) = +2,000,000$ lives.

We can also argue against expected utility maximization by considering the following case. Suppose that on every day n of eternity, with $n \geq 1$, you are offered the opportunity to pay half a unit of utility in exchange for playing a game with a $1/2^n$ chance of winning 2^n units of utility. By expected utility maximization, you would value the value of the game at $(1/2^n) \cdot (2^n) = 1$ units of utility, and at a price of $1/2$ units, it would be worth playing.

But consider what will almost surely happen if you adopt the policy of following expected utility maximization and playing the game, where "almost sure" is the technical term that probabilists use to describe an event that happens with probability one (such as getting heads at least once if you toss a fair coin infinitely many times). The sum of the probabilities of winning on the different days is finite: $1/2 + 1/4 + 1/8 + \cdots = 1 < \infty$. The Borel-Cantelli Lemma??ref then says that almost surely you will win only a finite number of times.[1] In other words, almost surely, there will come a day after which you will win no

[1] We can give an elementary proof of this fact in the case at hand (the proof generalizes to the general case). Let W_n be the event that you will win at least once after day n. Then $P(W_n) \leq 2^{-(n+1)} + 2^{-(n+2)} + \cdots = 2^{-n}$.

more. At that point, you may well be ahead, having won more than you paid. But the sum of what you won is finite, and from then on you will just lose half a unit of utility every day. Eventually, there will come a day when your losses will overtake your winnings, and from then on, you will just fall further and further behind every day.[2]

The very unhappy situation of playing infinitely many times and eventually starting to lose every time is the almost sure result of following expected utility maximization on each day. We can compare this to the neutral situation of refusing ever to play, and getting zero each day, or the situation of accepting the expected utility maximizing gamble for a number of days, until the probability of winning becomes really small, and refusing from then on.

It is worth noting that this is not just a paradox involving the aggregation of infinitely many utilities, except in the trivial sense that infinitely many zeroes make a zero (i.e., there is overall no benefit from playing the game once you stop winning). Almost surely, after a finite number of days, the expected utility maximizer falls behind the consistent refuser, and every day after that, the expected utility maximizer is further and further behind, like someone who got a subscription to a streaming service and forgot to either use or cancel it. And all these amounts are finite, and a finite, albeit unknown, distance into the future.

We can also consider an interpersonal version of the story. Suppose we have (countably) infinitely many people, numbered $1, 2, ...$, and person n is offered the chance to pay half a unit of utility in exchange for a chance $1/2^n$ of winning 2^n units. As before, by expected utility considerations it's worth it. So, if everyone is an expected utility maximizer,

---

Let W be the event that you win infinitely many times. Then $P(W) \leq P(W_n)$ for every n, since if you win infinitely many times, you must win on infinitely many days after day n, and so you must win on at least one day after day n. Since $P(W_n) \leq 2^{-n}$, we have $P(W) \leq 2^{-n}$ for every n. But probabilities cannot be negative, and the only non-negative real number x such that $x \leq 2^{-n}$ for every n is zero. So $P(W) = 0$, and hence almost surely W does not happen, so that almost surely you win only finitely many times.

[2]The example above uses exponential growth. More moderate growth will work, as long as the sum of the probabilities is finite. Thus, we could say that on day n the prize is $n(\log(n+2))^2$ and the probability of winning the prize is the reciprocal of that, since $\sum_{n=1}^{\infty} 1/(n(\log(n+2))^2) < \infty$.

everyone will pay. But by the Borel-Cantelli Lemma, almost surely, only finitely many people will win. Thus, almost surely, we will have infinitely many people pay a cost of half a unit each, and finitely many people win some finite amount. This is a disastrous situation, with a negative infinite overall utility. Almost surely, it would be much better if everyone refused to play, or only those who had a "non-negligible" chance at winning played.[3]

??refs:PrussBlog2011,Zhao, Wilkinson??ref:https://philpapers.org/rec/WILRAA-16

It appears that expected utility maximization cannot be rationally required. But it is the only clearly non-arbitrary solution to the problem of deciding under uncertainty.

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.[4] 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.

[3]It is worth noting that exponential growth is not necessary for the examples to work. All we need is that there is a chance p_n of winning a prize of $1/p_n$, and that $\sum_{n=1}^{\infty} p_n < \infty$. While for ease of calculation above I let $p_n = 1/2^n$, one can have much more moderate shrinkage, such as $p_n = 1/n^2$ or even $p_n = 1/(n(\log(n+2))^2$.

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

Expected utility utilitarians[5] will have a nice answer to this problem. But utilitarianism, as already noted??ref, has many highly counterintuitive implications.

????add: moral risk?

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 (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

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

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.[6] 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

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

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 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[7], 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.

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

A state presumably comes about when the people sufficiently agree to form a state, whose laws then typically need to be obeyed. But what constitutes such an agreement? Suppose that it is the agreement of a majority of those adults who make a reasonable effort to make their opinion heard (say, by voting). But where does the transition between a child and an adult lie? What effort is reasonable to require and what is unreasonable? And if we are to speak of the majority of the people, of *which* people? Presumably, the people inhabiting a given land. But there are many overlapping areas that can be considered the "given land", and in different overlapping areas majority opinions may be different. Moreover, what counts as inhabiting? (Suppose, for instance, someone lives different parts of the year in different places.) There is a vast multitude of questions to be answered by a majoritarian or any other account of the institution of a government, each question facing a Mersenne puzzle. And there are similar questions about the dissolution: Plausible, when a government becomes sufficiently unconcerned about the wellbeing of the people, it becomes illegitimate. But 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 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.[8] 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?

Political philosophy provides a number of examples. Consider the constitution problem. A state has a written or unwritten constitution specifying what must happen for legislation to be valid and hence authoritatively binding on the citizens. But how is a constitution instituted? One theory is that it happens by the consent of the people.??Aquinas But obviously for any state of sizeable size it will be false that all the people have consented: some have not made their opinions heard and some have been overruled. Requiring "consensus" or a supermajority raises the question of exactly how many dissenters can be tolerated, and once that question is answered we have a Mersenne question as to what grounds that cut-off being where it is. Requiring a simple majority or plurality involves one less free parameter: the cut-offs in having more than half of the voters or having more voters than any alternatives seem non-arbitrary. However, even a majority or plurality based system leaves questions about other parameters. Does one need a quorum of the governed? It would not seem right, for instance, to have a vote on a constitution on a day where the bulk of the population is unable to get the polls due to a hurricane or a war, and as a result only a small number of unrepresentative citizens can express their opinion. But if a quorum is required, then of course we have a Mersenne question as to exactly what constitutes quorum.

[8]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!).

Voting cut-offs and quorum are fairly easy to quantify. But what about the question of who the people giving their consent are? Presumably, small children should not be eligible. But where do we draw the line between small children and paradigmatic adult deciders? Any age-based line raises several Mersenne question—one about the numerical age cutoff and multiple questions about how age is measured (from fertilization, implantation, brain development, beginning of the birth process, completion of the birth process, etc.)? A cut-off based on mental capacity, on the other hand, involves many parameters that need to be set, because there is no single measure of mental capacity, and so one needs to have multiple measures with their respective weights. Moreover, we have a decision point on whether those who do not get a vote have proxies voting for them (e.g., parents) and, if so, who counts as whose proxy.

Or consider such details as how well-publicized the constitutional consultation needs to be, how clearly spelled-out the constitution needs to be for people's vote on it to be valid, and who gets to decide which options are presented to people?

Many of the above questions only make sense in the case of a formal consultation process of a sort that has occurred rather rarely over the course of human history. If we are not to think the vast majority of polities to be illegitimate, the account needs to allow for implicit consent, maybe of the sort involved in social customs. But there things become much less clear. There will almost always be some citizens who regard the state as illegitimate—indeed, some will regard any state as illegitimate. For many people, acceptance of a political system's legitimacy is not an simple binary question, but something that comes in degrees and has contexts. Imagine that two thirds of the population has a credence of two thirds that the political system is legitimate, and the remaining third of the population has a credence of ten percent in the system's legitimacy, and they express these credences in their actions. Should we then look at the average credence of the relevant citizens (e.g., those of age)—0.48 in my example above—and see if it meets some cut-off? If so, the cut-off will raise Mersenne questions. Moreover, people often do not just have a

single credence it the all-or-nothing legitimacy of the political system, but rather have different credences regarding different aspects of legitimacy: is this a state that has the right to use violence against its citizens to enforce laws, is it a state that has a right to levy taxes, to draft citizens to defend it or do other work for it, etc.

A set of Mersenne questions similar to those raised by the constitution problem is raised by the dissolution problem: the question of when it is that a political regime becomes illegitimate, and the respects in which it may be illegitimate (thus, perhaps, the traffic regulations of the Nazi state were legitimate).

One may wonder why questions about the legitimacy of a political system are being raised as part of a discussion of ethical questions. There are two reasons. First, we have a moral duty to obey the commands of a legitimate state. Second, only a legitimate state has the right to make certain kinds of onerous demands on its population. ??anarchism

Some readers will disagree with a number of the examples I gave. 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

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, moving 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.[9]

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,

[9]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. Social contract theories ground ethical rules in agreement between agents. We can divide this based on whether the agreement is actual or hypothetical. Actual agreement theories face obvious problems. First, it is highly implausible to think of the typical agent in society as having *actually* agreed to live by moral rules, apart from special cases such as a pious person vowing to God to sin no more. Second, actual agents can agree to live by unjust rules, even rules unjust to themselves, and such rules would not constitute morality.

Contemporary social contract theories tend instead to be 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.

The risk and uncertainty point is worth emphasizing. Some hypothetical agreement theorists think that rational agents would only agree to rules that do not treat anyone inhumanly.??refs But a rational agent who is more accepting of risk will be willing to tolerate rules that create a minority group that is treated inhumanly if the risk of being a member of that group is sufficiently small—i.e., if the group is a small enough fraction of the general population—and the the benefits to the majority are sufficiently large. There will be types of inhuman treatment and levels of risk that it would not be rational to accept for the sake of a high probability of a large benefit, but the lines between these and the ones that it would be rational to accept do not seem derivable from any plausible set of basic principles of rationality.

Granted, typical Kantian constructivists will insist that certain kinds of inhuman treatment would never be rationally acceptable. But now consider the Mersenne questions about these kinds of treatment. For instance, destroying the autonomy of another person might be taken never to be rationally acceptable. But a minor limitation on another's autonomy clearly is acceptable for a sufficiently great good: if the only way to save a country from nuclear destruction by evil enemy would be to acquiesce in the enemy's demand that everyone wear jeans on Friday, then this limitation on sartorial autonomy should be enforced. Somewhere there is a line between minor limitations of autonomy and such deep destruction of autonomy that could not be tolerated no matter the price. The only "natural" place to draw this line would be at *complete* destruction of autonomy. But if it is only complete destruction of autonomy that is prohibited by the Kantian, then this does not place a sufficient constraint on the rules that could be accepted. For instance, the enslavement of persons would not be prohibited, as long as the enslaved persons were still capable of some autonomous agency, no matter how minor.

Furthermore, even prohibiting treatment that completely annuls someone's autonomy will not avoid Mersenne questions in the vicinity. For we will have probabilistic questions. Is it permissible to perform an action that has a 99.9lines to draw seem to be at 100autonomy prohibits nothing: any action we perform can fail. A prohibition on an action that has a 0prohibits everything: I scratch my head, and there is a tiny chance that due to some weird sequence of events this causes an earthquake that leads to you getting hit on the head by a beam that results in your life being reduced to a vegetative level. And while 50much more reasonable, in some difficult cases it is excessively restrictive. If a child is certain to die within a day, and is suffering from horrific pain that can only be relieved by a drug that has a 50of the day, administering the drug can be permissible.

5.5. Positive law models. ????

5.6. 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.

5.7. 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.

An initial 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.

Suppose the initial concern can be allayed in both of its forms, perhaps by clever development of a theory that does generate the vague moral claims and by biting the bullet and admitting that moral agonizing is out of place in these vagueness cases. There is still another question: how do we account for the vagueness here. There are three main contemporary accounts of vagueness: (a) non-classical logic, (b) supervaluationism and (c) epistemicism.

On non-classical logic approaches to vagueness, one typically increases the number of truth values beyond two. Consider an ethical Sorites series, where we fix some circumstances C and then say:

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

Now for each positive integer n, the following material conditional sounds plausible:

(A_n) If giving \$1000 to a stranger is better than giving \$$n$ to one's parent, then giving \$1000 to a stranger is better than giving \$$(n+1)$ to one's parent.

From A_0 and A_1, one concludes by *modus ponens* that giving \$1000 to a stranger is better than giving \$1 to one's parent. From this and A_2, by *modus ponens* 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. Thus, we need to reject one of the premises A_n. Presumably it's one with $n > 0$, since A_0 is clearly true. But a material conditional $p \to q$ is false just in case p is true and q is false. Hence, if A_n is false for $n > 0$, we have:

(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.

And that is exactly the kind of sharp transition that the vagueness theorist wishes to deny.

The non-classical approach to vagueness typically involves a logic with many truth values, e.g., a truth value for every number between 0 (fully false) and 1 (fully true). Then 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. If we have a large enough number of truth values, we can accept this at face value.

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. 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.

Finally, the application of supervaluationism to ethics is itself deeply problematic. 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 a realist 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, rather than the commands found in some other location.

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 neither that theory nor the empirical data on the patterns of language use are available to us in sufficient detail to settle the meaning of vague terms. 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.

But the problem here is exactly the same as the last problem with supervaluationism. Ethical questions are turned into purely verbal questions. Just as on supervaluationism, there is a multiplicity of concepts closely corresponding to our words "right", "better" and "reason". On supervaluationism, none of these concepts was privileged, which turned ethical questions into purely verbal questions, undercutting the idea that our lives are to be governed by reasons and morals. On epistemicism, there *are* privileged concepts that exactly correspond to the words, but they are privileged purely linguistically—it just so happens that these privileged concepts better fit with our usage under the correct semantic theory. We get 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.

In summary, non-classical logic violates classical logic, which should only be a last resort, and further multiplies rather than resolving Mersenne questions. Supervaluationism likewise multiplies Mersenne questions. And, perhaps most seriously, both supervaluationism and epistemicism as applied to ethics turn ethical questions into purely verbal ones, undercutting a robust realism.

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.

7. Hume's objection: Complexity, instinct and nature

Hume saw the complexity of property, inheritance, contract and jurisdiction, and used this complexity to argue for apparently *against* a Natural Law account:

> For when a definition of *property* is required, that relation is found to resolve itself into any possession acquired by occupation, by industry, by prescription, by inheritance, by contract, &c. Can we think that nature, by an original instinct, instructs us in all these methods of acquisition? (??Enquiry::Morals)

To further expand on the complexity, Hume notes the vast variety between these rules in different societies, and analogizes them to the variety of architectures found in housing across societies. A better account, Hume insists, is that we simply engineer rules for the

sake of social utility, just as we engineer houses for various ends, and in different environments this results in different solutions, albeit ones with a lot of commonality.[10]

Three responses are possible.

First, we do actually have good empirical reason to think that our moral intuitions and instincts vary quite a bit from case to case. An account of our actual moral instincts is likely to have enormous complexity. Granted, some of the complexity of our actual moral instincts is due to failures. For instance, racist or sexist biases introduce complexity by distinguishing cases that do not morally differ. And it would be *correctly functioning* instincts that we would expect the natural law to be expressed through. Thus, even if Hume's argument fails with respect to our actual instincts, it may work with respect to correctly functioning instinct, since we have reason to think that this would in some respects be simpler than our actual instinct.

However, in fact, it is far from clear that purifying our instinct of malfunctions would make for so much simplicity as would be needed to support Hume's argument. Removal of some biases will indeed remove some complexity. But enough complexity is likely to remain that Hume's argument will not be convincing. Moreover, it is likely that our instincts sometimes fail through not making distinctions that should be made, and hence in some respects our correctly functioning intuitions would likely be more complex.

Second, the account I am defending is one on which our forms set norms. These norms can be arbitrarily complex. Granted, there will be a harmony between these forms and our instincts and intuitions. But this harmony is not a one-to-one mapping. There is such a thing as normal and abnormal food for a type of organism, and we expect the organism to have instincts that tend to direct them to normal consumption and away from abnormal consumpton. But just as correctly functioning sight can still err, so too a correctly functioning feeding instinct can lead organisms to ingest what is abnormal and to refrain from what is normal, especially in environments that have abundances different from the

[10]It is natural to try to solve the problem by adverting to positive law. But Hume notes that there is much complexity with respect to the institutions by which positive laws are produced.

ones present where the organisms evolved. Thus, our natural instinctive preference for high-calorie foods leads humans in affluent Western countries to abnormal consumption, and hence to Reflection can gain additional information as to normative consumption on the basis of high rates of obesity.??refs By reflecting on our nutritive instincts and *other data*—such as the teleology of nutrition and medical facts about us—we can get additional information as to what is normative consumption for an organism, including a human one. This additional information is still fallible, and may fall short of the complexity of the norms involved. And what goes for our nutritive instincts is even more strongly applicable to our moral ones.

Third, recall Hume's own solution to the problem that the complexity of the rules is a result of our social engineering for social utility. Hume's solution is subject to complexity problems of its own. For instance, what groups count as societies and how do we aggregate the benefits to individuals to get a social utility, etc.? But something similar to Hume's solution can be appropriated for the natural law account. We can suppose norms in our nature establishing the goals for certain social institutions, such as property or state authority, and perhaps establishing some constraints on how these goals are to be pursued, and at the same time requiring us to engineer institutions that satisfactorily pursue these goals within the contraints. These norms will be complex, but will be less complex than the vast complexity in our social institutions. Thus, we have a bootstrapping, from fairly complex norms setting the ends and constraints for social institutions, to the institutions themselves.

CHAPTER III

Ethics and metaethics

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. What are moral or rational norms?

The idea that norms are species relative suggests that in the space of possibilities—and perhaps in extraterrestrial reality as well—there will be species of beings that are intelligent enough to have advanced science and technology, but whose natural behavior will be quite different from us. This raises a difficult question as to what makes a particular set of natural norms count as a set of moral or even rational norms, and hence makes the species that possesses them a species of moral or rational agents.

Not all natural norms are moral or rational norms. The natural norm behind a properly functioning horse shedding in the spring is neither moral nor rational. Plausibly, a necessary condition for a moral norm is that it govern voluntary behavior. But the question of

what behavior counts as voluntary is difficult. It is tempting to say that the behavior of an entity is voluntary if it is subject to reasons. But reasons live in a space made possible by rational norms, and so it seems we need an account of rational norms, at least, to make sense of what behavior is voluntary.

Here is one highly speculative Aristotelian functionalist way to answer the questions. First, we connect reasons and norms with goods considered as such. An (internal) reason for a behavior is a representation of the behavior as *good*. A mouse may represent cheese as yummy, or maybe even as nutritious, but not as *good*. Of course, being yummy or being nutritious is thereby good, but to represent as yummy or nutritious is not to represent as good.

Then a necessary condition for a behavior to be voluntary is that it comes from such a reason. This, of course, raises the infamous problem of in-the-right way. A behavior can be caused by a reason without being voluntary. The famous case is the belayer who intends to murder a climber by dropping the rope, and then his hands start shaking at what he has intended to do, which results in an involuntary dropping. Whatever reason he had for the murder is the cause of the dropping, but the dropping is involuntary. Aristotelian metaphysics does, however, seem to have a tool for solving this problem. Causation can be seen to be teleological in nature, and we might say that it is a primitive fact that sometimes the effect *is* a fulfilment the teleology of the cause, in which case we can say that the effect is caused in the right way. A voluntary behavior is one which is a fulfillment of the teleology of the cause.

Finally, the will can then be functionally defined as the system by which reasons lead to behavior. A rational norm is a norm of behavior of the will favoring some or all reasons, and a moral norm is a norm of behavior of the will favoring some or all reasons that themselves are focused on a good not considered primarily as a good to self.

This is not, of course, the only way to define which norms are moral or rational. And it is quite possible that the question of which norms are moral or rational is largely a verbal question. Go back to the characterization of reasons in terms of goods. The mouse takes the

cheese to be yummy, and that is not taking the cheese to be good. But humans often represent good things in thicker ways: as beautiful, courageous, or even divine. Couldn't we imagine a continuum of animals where at one end the cheese is represented as yummy and on the other as having gustatory beauty? Somewhere in the continuum we have moved to representing the cheese as having a thick form of goodness. But where this happens is unclear.

We have a certain set of norms for the will. We can call these rational, and a subset of them moral. What hangs on whether a different set of norms governing the behavior of a different class of beings counts as rational or moral? Couldn't this be like the question of which animals' hard projections count as horns?

But perhaps the question of which things are moral agents matters for first order moral questions about interspecies relations, such as which organisms we are permitted to eat, whose lives we should save, etc. However, it is not clear that the answers to these questions will neatly line up with determinations of the boundaries of moral agency. Consider intelligent whales that are fine philosophers and scientists in their epistemic life, and that even enjoy contemplating the good, but do so purely non-practically, without the good being any motivator for their actions, which are all instinctive. It would seem wrong to each such beings, even if they turn out not to be moral agents. On the other hand, imagine a shrimp that has the normal behavioral complement of a shrimp, with one exception: it represents the ingestion of algae as good, and voluntarily pursues this good as such. And its intellectual abilities are the minimum needed for an ordinary shrimp's life as combined with the most minimal possible concept of the good. It may well be wrong to eat such shrimp, but given a choice whether to save the life of one of these minimally moral shrimp or the life of one of the intelligent but amoral whales (of course, we should not be biased here by size!)

3. 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.

4. Supererogation

??cut?

5. Supervenience

It is widely held that moral facts supervene on non-moral facts: if two possible worlds differ with respect to moral facts, they must differ with respect to at least one non-moral

fact. Similarly, it is held that normative facts in general supervene on non-normative facts. The difficulty is then to explain the supervenience relations.

The nature-first theorist has a complex relationship to both supervenience claims. Let us begin with the supervenience of the normative on the non-normative, and first consider general normative claims such as that every sheep should have four legs and every human should refrain from torturing the innocent. Such normative claims are necessary truths, since their truth is a part of what makes a sheep a sheep and a human and a human. Necessary truths vacuously supervene on any basis we might choose, since there are no possible worlds that differ with respect to necessary truths.

But what about *particular* normative claims, such as that Sally ought to have four legs or that Biden ought to discharge the duties of the President of the United States? If we are interested in the supervenience of the normative on the non-normative, we face a serious problem on Aristotelian metaphysics: there are very few non-normative facts. It seems that every natural kind is defined in part by normative properties. That Sally is a sheep is itself a normative thesis, since a part of what it is to be a sheep is to be such that one ought to have four legs. That Sally is an animal is also normative. And Sally is *essentially* a sheep, and hence being a sheep is central to Sally's identity in such a way that it may even be the case that even the claim that Sally exists may count as a normative claim. Aristotelian metaphysics likely reaches even down to the fundamental particles. Electrons not only do but should repel other electrons. A non-normative fact, thus, will not make reference to any natural kinds, and not even to any particulars falling under natural kinds.

On Aristotelianism, every particular, with the exception of God if there is a God, falls under a natural kind. But facts about God are through-and-through normative, since God is essentially perfectly good. Thus, on Aristotelianism, all facts about particulars are normative. Moreover, on Aristotelianism, a non-normative fact cannot include anything existential. For to be is to be a substance or to be appropriately related to a substance (say, by being its accident). And a part of what it is to be a substance is to have a form that specifies how one should behave. Thus, what it is to be is in part to have norms or to be related

to something that has norms. If so, then every existentially quantified claim is normative. The denial of a normative claim is normative as well, and since a universally quantified claim is the denial of an existentially quantified claim (everything is *F* if and only if there does not exist an object that is not *F*), universally quantified claims will be normative as well.

But if all facts about particulars and all quantified facts are normative, it seems that *all* facts are normative on Aristotelianism. If this is right, then to say that two worlds are the same in non-normative terms is to say literally nothing about them. And if all facts are normative, then any two worlds that are the same in normative terms are altogether the same. Thus, the thesis of the supervenience of the normative on the non-normative becomes the thesis of modal fatalism: that there is only one possible world! And we have good reason to reject this thesis.

But while this goes against mainstream views of normativity, it is arguably an advantage of the view. For by eliminating non-normative facts, we no longer have any puzzling phenomenon of the relationship between the normative and the non-normative to be explained.

What about the moral supervening on the non-moral? Moral facts are normative facts about the will. Again, general moral facts, such as that humans should refrain from torturing the innocent or should discharge the duties of the President of the United States if they have voluntarily sworn the relevant oath of office, are necessary truths and hence trivially supervene on whatever facts we want, including non-moral or even non-normative ones. But there are many particular normative facts, such as that Biden should discharge presidential duties, that depend on facts about human wills, such as that Biden *voluntarily* swore the oath of office, and since it is the very nature of the human will to be such that various moral facts about it hold, facts about human wills are not going to be among the non-moral facts. Thus the Aristotelian will also reject the supervenience of the moral on the non-moral.

But what the intuition underlying the supervenience claims? We can put the intuition as follows. Imagine a world that *looks like* ours. It has bipeds that look just like us. There is, for instance, a biped that is empirically indistinguishable from Biden. The history of these bipeds is empirically indistinguishable from our history. Could it really be that the moral facts about these bipeds be other than about us? Could it be, for instance, that although these bipeds behave just as we do towards those who torture people at random, among them there is nothing morally wrong with such torture?

6. Outlandish paradoxes

It is easy to generate paradoxes in ethics and decision theory by invoking outlandish situations. Many such situations involves infinities. I will give two representative examples.

First, we have the Satan's Apple paradox about infinite sequences of choices on which something further depends:

> Satan has cut a delicious apple into infinitely many pieces, labeled by the natural numbers. Eve may take whichever pieces she chooses. If she takes merely finitely many of the pieces, then she suffers no penalty. But if she takes infinitely many of the pieces, then she is expelled from the Garden for her greed. Either way, she gets to eat whatever pieces she has taken. ??ref

The puzzle is that for each piece, Eve has conclusive reason to take the piece, but if she acts on all these reasons, something terrible happens. As presented, this is a paradox about self-interest, but we can turn it into an ethical one by supposing that the rewards and penalties of Eve's choices devolve on someone else, say Adam. In that case, we can say that Eve should accept each piece and yet that's the worst option.

Another kind of paradox involves infinite numbers of beneficiaries. Imagine that there is an infinite number of complete strangers, numbered with the integers (negative, zero

and positive), as well as two cats, all facing a deadly danger, and you have a choice between one of three equally convenient options:

(10) Save the strangers numbered $0, 1, 2, ...$

(11) Save the strangers numbered $-1, -2, -3, ...$ and one cat.

(12) Save the strangers numbered $1, 2, 3, ...$ and two cats.

Now, you have no reason to prefer the stranger numbered 0 over the stranger numbered -1, the stranger numbered 1 over the stranger numbered -2, and so on. So as far as the saving of people, (10) and (11) are a wash, but it's better to save a cat than not to, so (11) is morally preferable.[1]

But likewise there is no reason to prefer saving the people numbered with negative integers over the people numbered with positive integers, so as far as the saving of people goes, (11) and (12) are balanced. However, saving two cats is better than saving one, so (12) is better than (11).

But now, (10) is clearly better than (12): for in (10), you save stranger 0 instead of the two cats, and wonderful as cats are, it is much better to save that one human over two cats.

So we have a moral preferability circle, and whatever you do, there is something better you could have done at no greater cost. It seems plausible that you have a duty better if you can do so at no greater cost, and yet whatever you do, you violate that duty. And so it seems that you cannot act as you ought, thereby violating the plausible maxim that ought implies can.

There have been various attempts to defuse such paradoxes,???refs and a defender of human nature as the foundation of ethics can accept any of them. However, there is also a simple and highly intuitive alternative to these defusions. A horse's nature may ground facts about the appropriate gait when browsing on grass and the appropriate gait when fleeing a predator through water. But equine nature is simply silent on a horse's gait when fleeing aliens in a zero-gravity environment. Similarly, our human nature could be silent

[1] If the reader thinks that cats do not fall in our moral purview, just replace the saving of a cat with saving a human from some minor harm.

on how we should act in outlandish situations, and our principles just need not extend to such cases. This fits very well with the ordinary person's disdain for philosophers (like me) who spend a lot of time thinking about such cases.

There is another similar solution. It could be that our ordinary moral rules *do* extend to outlandish cases. Thus, the moral reasoning by which we generated the moral paradoxes in Satan's Apple and the infinite saving case may be correctly grounded in norms in our nature. It may well be that, say, (11) is morally better than (10), that (12) is morally better than (11), that (10) is morally better than (12), and that you ought to do the morally best (or one of the morally best, if there is a tie) between the three options. It's just that these specifications of our nature are impossible to fulfill under these circumstances. In other words, it is very plausible to say that ought implies can in situations that are a part of humans' natural environment, but there may be logically possible outlandish situations that go far beyond this environment where ought no longer implies can. Insofar as our nature gives us norms fitted to our human environment, we should not be surprised if these norms have counterintuitive implications, such as violating ought implies can, in situations far outside that environment.

Similar solutions will be available on at least two other ethical theories where the laws may be customized to humanity: contractarianism and divine command theory. But, on the other hand, such solutions will be implausible on theories that purport to apply to any kind of rational being at all, theories such as utilitarianism or Kantianism.

A similar point can be made about outlandish epistemological paradoxes. ??refs-and-examples On a natural law epistemology, we should not expect our nature to give us guidance, or at least satisfactory guidance, in situations too far out of the human environment. And while in ethics there are at least two common anthropocentric alternatives to natural law, contractarianism and divine command, in epistemology anthropocentric alternatives are harder to find.??Hawthorne? Thus, we have perhaps an even stronger consideration in favor of a normativity based on human nature on the epistemological side.

More will be said in ??forward about outlandish scenarios.

7. Agent-centrism

7.1. The egoism objection. According to Natural Law metaethics, an action is right provided that its performance constitutes the will's flourishing, and is wrong provided that its performance constitutes the will's languishing. This seems objectionably egoistic. Paradigm cases of moral wrongness involve harm to others, and are wrong because of that harm. The thought that the action makes the agent languish is a thought too many.??refs Therefore, the argument goes, we should opt for an other-centered metaethics.

My response will be two-pronged. First, I will argue that the very features criticized in Natural Law provide a significant advantage in a number of cases. Second, however, I will argue that the argument against Natural Law's agent-centric character only works against some normative developments of the Natural Law metaethics, rather than against the metaethics.

7.2. The normative advantages of agent-centrism. Other-centered theories nicely account for what is wrong with murder: it gravely harms the victim. They account slightly less well for what is wrong with typical cases of attempted murder: it is an attempt to harm to harm the victim. But they do not account for atypical cases of attempted murder where the victim simply does not exist. Suppose Alice thinks that she has an identical twin living somewhere in Toronto, and sets out to kill her, to avoid the twin's claiming an inheritance. Alice has an extremely rare genetic disorder which an identical twin would share, but which is very unlikely to be otherwise exhibited even in a city as large as Toronto. She adds a poison to Toronto's water supply that targets only people with this genetic disorder, and then takes care to avoid drinking Toronto's tap water. But in fact Alice never had a twin.

There is thus no one that Alice is attempting to kill. Yet morally speaking, she is just as guilty as in an attempted murder case where her twin exists, and depending on one's views on moral luck maybe even as guilty as in the case where she succeeds in killing her twin. It is worth noting that in Anglo American jurisprudence, Alice might get away under

the doctrine of impossible attempts, on which an attempt has to have some feasibility, and trying to kill a non-existent person has none. (Of course, Alice is likely to be convincted for pollution and for reckless endangerment of people with this disorder, but these are lesser evils.) However, it is clear that notwithstanding the law of impossible attempts, it makes no difference to Alice's guilt whether in fact she has a twin or not.

We may, of course, try to save the doctrine that wrongs are always wrongs to another by trying to identify other victims, such as society or God. We can try to tweak the Alice case to exclude the society solution. Perhaps Alice is trying to kill her twin sister in a world where she thinks they are the only survivors of a disaster, but in fact Alice is the only survivor and she's never had a twin. That won't help with the God case, at least not within classical theism, since God is traditionally thought of as a necessary being??Refs. Furthermore, the intuition that I am responding to is that there is something particularly centered on the ordinary direct human victim of a wrongdoing that contributes much of the wrong. And if God or society is what we count as the victim in the Alice case, then it seems that we have to say that in the Alice case there is less wrong than in the more ordinary case of attempted murder where the victim actually exists. And yet it seems that how wrong Alice's action is does not depend on whether she has a twin.

The above focused on patient-centric wrongs. We can also think about patient-centric duties. Again, it seems that our duties go beyond these. Plausibly, we have a duty not to deliberately produce a human being who is so genetically constituted as to be practically guaranteed to have a life of unremitting suffering. Now suppose that Bob and Carl both know that they and their spouses have genes such that if they reproduce, the child will have a life of unremitting suffering. In light of this knowledge, Bob refrains from reproduction, while Carl's sadistic tendencies impel him to reproduce in light of this fact. Bob and Carl both have a duty. In Carl's case, we can identify the individual to whom he has this duty, an individual that that he has wronged. But in Bob's case, the analogous individual does not exist—precisely because Bob has fulfilled his duty. Again, we can try to identify others to whom Bob owes not having a child—society, God and likely Bob's wife.

But since there is one less individual here than in Carl's case, Carl has a somewhat more stringent duty, since the unfortunate child exists. However, it does not seem that Carl has a more stringent duty than Bob.

Finally, moving away from cases, it is obvious that some degree of agent-centrism is needed for any plausible story about wrongs and duties. It is *agents* that do wrongs and have duties. We have a duty not to eat humans. Lions, pigs and horses have no such duty. Therefore, an account of what makes it wrong for me to eat other humans has to involve some facts about *me*. It cannot be wholly other-based.

One might respond that on the Natural Law account, what makes it be wrong for me to eat other humans involves *only* facts about me, while it should also involve facts about the prospective victims. But how we understand this objection depends on how we read questio of "what makes it wrong for me to eat other humans".

First, we can understand it as a question about the grounds of the general moral rule that it is wrong for humans to eat other humans. On Natural Law the grounds of that general moral rule are entirely within the agent. However, that is how it should be. For the general moral rule would also hold even if there were no other human beings in the world. And in fact the general moral rule would hold *non-trivially* even if there were no other humans, since even if there were in fact no other human beings, I should avoid actions that are likely to constitute the eating of a fellow human being (e.g., shooting and eating an animal that has a significant epistemic probability of being human). An account of the wrongness of eating other human beings that requires other humans to exist is unsatisfactory.

Second, we can understand the question as asking about particular cases: What is it that makes it wrong for me to eat, say, Carl? But then the Natural Law story is going to include a fact about Carl: I am the sort of thing that shouldn't eat other humans and Carl is another human.

On neither reading do we have an argument against Natural Law metaethics.

7.3. Avoiding agent-centrism in normative Natural Law ethics. Here let me start with a personal confession. For many years I objected to the eudaimonism I took to be at the heart of Natural Law, which one might take to consist of the twin theses:

(13) What makes an action right is its promotion of the agent's flourishing.

(14) An agent's right actions are aimed at the agent's flourishing.

And these theses seemed objectionably egoistic.

However, while there are ways of pairing the Natural Law metaethics that I have been developing with a normative ethics that embraces (13) and (14), they are both dispensable.

Indeed, (13) is so clearly wrong that it is unlikely that many Natural Law theorists accept it, given the well-known anti-consequentialism of the Natural Law community. It is wrong to rob a bank in order to pay for the tuition of an ethics class even if there is strong empirical evidence that this class will be so transformative that on the whole one's flourishing will be promoted, even if one takes into account the temporary harm done to it by the robbery. Similarly, one may have a duty to continue working in a job that is just barely moral and empirically likely to be destructive of one's flourishing as a moral agent in order to pay the medical bills for a child.

Now, on the metaethics that I am defending, what makes an action right is that it *constitutes* the agent's flourishing with respect to the will. If we add to this the thesis that whatever constitutes the agent's flourishing with respect to the will constitutes the agent's flourishing as a whole, then we get a version of (13) with "promotion" replaced by "constitution". However we should not think that what constitutes the agent's flourishing with respect to the will constitutes the agent's flourishing as a whole. If an agent is flourishing with respect to the will, but is full of ignorance, in great pain, and lying in a bed of vomit??Vlastos-ref, the agent is not flourishing on the whole.

And the thesis that what makes an action right is its constitution of the agent as flourishing in respect of the will seems to be simply a thesis about proper function, and does not imply any selfishness. Consider, for instance, that a bee's defending the hive at the expense of its life fulfills the bee with respect to whatever we call the driver of the bee's

activity (we may not wish to call it a "will"). But this does not make the bee in any real way selfish. And certainly a guided missile is not selfish just because it fulfills its nature by exploding.

It is tempting to think that in the case of an agent who is driven by a rational will, if what makes the action right is its constituting the agent as flourishing (in one respect), then by willing the action under the description "right action", the agent aims at flourishing, in a way in which neither the bee nor the guided missile's actions are aimed at flourishing. If this line of thought is correct, then the characterization of rightness in terms of flourishing implies (14), and that seems more objectionably egoistic.

However, a rational agent's intentions are hyperintensional. It is possible to aim at heating up a room without aiming at increasing the kinetic energy of the molecules in the room, even though what makes there be heat in a room is the kinetic energy of molecules, and necessarily one is present if and only if the other is. Indeed, during the millenia before the relationship between heat and kinetic energy was known, no one aimed at increasing the kinetic energy of molecules while heating a room, and even now when the relationship is well-known, few people's intentions in turning a thermostat make reference to molecular motion. Similarly, even if the rightness of an action is grounded in, constituted by or even identical with the action's being an instance of the agent's flourishing as a willer, the rightness can be aimed at without aiming at the flourishing.

Furthermore, as has often been pointed out in the literature??(on moral fetishism), virtuous agents rarely aim at rightness as such. Instead, they aim at thicker right-making features of an action, such as its being an expression of loyalty to a friend, its fulfilling a stranger's need, or its having been promised. It is because the action has such thick features that its performance is an instance of volitional flourishing.

There are, of course, times when a human agent aims at rightness as such. One set of cases is provided by agents who cannot figure out on their own what is right and have to take the rightness of an action on the authority of another, without understanding what makes the action right. Such agents include small children, but also sometimes ordinary

well-functioning adults who find themselves in such situations of such moral complexity that they turn to a professional ethicist or a trustworthy friend for advice. Is this objectionably egoistic, assuming the rightness is constituted by the agent's volitional flourishing? There are at least three reasons to doubt this. The first was already mentioned: the hyperintensionality in intentions.

To see the second reason, observe that we actually have a *three layer* story:

(15) the rightness of the action

(16) the action's being such as to constitutive volitional flourishing

(17) the thick features of the action because of which the action constitutes volitional flourishing.

The egoism objection in the case of agents aiming at right as such insists that the willing of (15) inherits an egoistic character from the agent-centrism of (16). But note that (16) is not the end of the story. Just as (15) is grounded in (16), so likewise (16) is grounded in (17). And in paradigmatic cases, the thick features in (17) are other-centric features. If we think that willing the rightness of the action inherits egocentrism from the flourishing, we should even more think that it inherits other-centrism from the thick features, since the thick features are a yet more ultimate ground of rightness than the flourishing is.

Finally, recall that we already noticed that an action can be right and constitute volitional flourishing but hamper one's flourishing as a whole, as in the case of working a soul-destroying job to pay family medical bills or refusing to rob a bank in order to pay for a morally transformative class. In such cases, it is absurd to say that by aiming at volitional flourishing one is being selfish, since the action does not, in fact, contribute to one's good overall.

Indeed, a metaethics that grounds rightness in flourishing as a willer is compatible with one's never being required to intentionally pursuing one's own good. We can (perhaps with some difficulty) imagine an alien species whose members pair off in such a way that the proper functioning of each one's will is just to will the good of the other member of

the pair. Perhaps this is a species so physically constituted that they are always more effective at benefiting others than at self. In such a species, the Natural Law metaethics would require utter unselfishness—and yet what would *ground* the rightness of an action would be that the action is proper to one's will and hence constitutes the will as flourishing.

We could imagine two versions of such aliens. They might be less reflective than ourselves, and never act on higher-order reasons like rightness as such. Or they might be reflective, and might even come to a Natural Law metaethics on which an action is made right by its constituting the agent as flourishing. In such a case, they might aim at an action under the description "right", but only because they know that an action's rightness is ultimately grounded in their species in the action's benefiting the other member of the pair, though mediately in its constituting the agent's flourishing.

In the latter case, it is not completely clear that such aim at rightness counts as *intending* rightness. For the rightness is neither an end in itself nor a means to benefiting the other. We might say that here we have a "calibrational aiming" at rightness rather than an "intentional aiming" at it. For an analogy, we might consider Donnellan's famous example of being told that the man drinking champagne is a senator??ref, where in fact the man is drinking gingerale. Now imagine an assassin who uses this information to aim her gun at the senator and kill him. The assassin is calibrationally aiming at shooting a champagne drinker, but does not care, either as a means or as an end, whether the person she kills is drinking champagne (she's not, we may suppose, a temperance activist!). Consequently, if her bullet kills the senator, her wicked action is successful, even though she did not kill a champagne drinker. Or so one might think. And if one thinks this, then one might be able to say that even in the self-reflective variant of our altruistic aliens, there is no intention to do the right thing—there is only calibrational aim at the right thing. And if we say that, then there is even a possibility that the objectors to moral fetishism(??explain) are right, and even we humans should not aim at rightness. Instead, we should aim at the constituents of rightness, but sometimes we do so by calibrationally aiming at rightness.

I do not endorse this possibility, but it is available to those who want to stay away even from that modicum of self-centeredness involved in intending rightness.

CHAPTER IV

Applied ethics

1. Introduction

Thinking that ethical duties are grounded in norms innate to human nature does not by itself logically entail answers to controversial questions of applied ethics. One can think that our nature requires us to kill those whose suffering we cannot stop, and hence that euthanasia is required, one can think that our nature prohibits the killing of the innocent even if that killing would be in their interest, and one can have an in-between view.

But the nature-based approach provides at least two benefits for applied ethics. First, because of the Aristotelian harmony principle, it allows facts about our natural behaviors and needs as the kinds of organisms we are to to provide us with defeasible but often strong evidence about what we should do. Second, it makes it more plausible than it would be on a number of competing theories that the answers to applied ethics questions might be irreducibly intricate—not reducible to a small number of simple principles—and might include domain-specific ethical rules for the various areas of our natural lives, such as family relationships or sexuality or (if it's natural) property rights.

We will thus explore some things that we can say on nature-based ethics. The plausibility of what we will say will serve as indirect evidence for the underlying Aristotelian metaphysics.

2. Natural relationships

2.1. Siblings and cousins. An interesting test case for an ethical theory is whether it 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: such rules can be fundamental. And the laws can, in principle, be at any level of precision, be it to simply consider one's siblings at a higher weight in one's moral calculus than more distant relatives (we are all relatives, after all, as we learn from evolutionary theory) or to prefer one's siblings over one's cousins to such-and-such a specific degree. The laws could even have social construction built in: they could require us to respect our relatives in the ways that our society prescribes, and require us to establish societies that institute ways for us to respect our relatives.

Divine command theory has a similar advantage: God's commands can be at any level of generality or precision, be it to love one's neighbor or to telephone one's cousins at least twice a year if one can. In principle, rule utilitarianism can do this as well: it is plausible that having rules concerning special relationships like fraternal ones could maximize utility. But Natural Law arguably gives a better explanation of the duties tied to these special relationships. For the nature of these special relationships is very plausibly tied to our humanity, and hence it makes sense that the special obligations attached to them should flow from that humanity rather than the commands of a God or the results of an abstract hypothetical optimization procedure.

Indeed, on Aristotelian natural law, we can say that having these kinds of special obligations is an important aspect of what makes us human—for it is an important aspect of our form, which is precisely what makes us human.

2.2. Less natural relationships. We have a broad variety of socially-instituted and culturally-variable relationships which are very unlikely to have norms encoded for them in human nature. In English-speaking countries the relationship to the parent of one's godchild or the godparent of one's child tends not to have sufficient importance to even have a name, while in other cultures it is important and specifically named. The relationship between an employer and employee varies so broadly with legal and social structures

that it is probably best seen as an umbrella for a number of different relationships, none of which is likely to be encoded in human nature.

Admittedly, a relationship could fail to be culturally widespread and yet could have norms encoded for it in human nature, but there is a more elegant approach to analyzing such relationship: we can see them as cultural determinations of a more fundamental relationship type, with some of the norms coming from human nature's rules for the more fundamental relationship and others from the culture. Moreover, human nature may prescribe the scope for cultures to establish the rules. Such relationships can be thought of as "less natural". At the same time, the difference between these relationships and the "more natural" ones like siblinghood are likely to be largely of degree. For while there may be a fundamental normative relationship of siblinghood, it has further culturally-determined norms.

2.3. Marriage. A particularly interesting question, of significant relevance to controversies in our society over the past century, is where *marriage* lies on the naturalness spectrum. I shall argue that it is likely to be quite natural, with a number of fundamental norms grounded in our human nature by arguing against two main alternative theories and combinations of them.

The first theory holds that marriage is an institution defined by many human societies. Like other such social institutions, such as judgeship, parliament membership, monarchical sovereignty, exchequer chancellorship, and presidency, it is defined by the rights and obligations conferred by society on those who enter into the institution. While we use the same words "judge" and "monarch" across societies, there is only a family resemblance between the institutions these terms refer to, since the actual rights and obligations defining the institutions are often very different indeed. The resemblance may be very weak: the rights and obligations of the monarch of England in the 13th century are about as different from the rights and obligations of the current monarch as the rights and obligations of modern day judge are from a modern day executioner. Nonetheless, for historical

reasons we may use the same word "monarch", sometimes clarifying with adjectives like "absolute" or "constitutional".

The second theory has it that couples choose to undertake certain obligations with respect to each other, which obligations give rise to rights, and this complex of rights and obligations defines the marriage. In more traditional societies, a couple may not choose the obligations specifically but rather will simply opt for the "customary" obligations and their consequent rights. In modern Western society, many couples write their own wedding vows, specifying general obligations. But even in those cases, it is likely that these vows are not typically thought of as a precise and exhaustive legal contract, but rather as a way of customizing one of the prevalent packages of obligations. Again, on the individual theory, we use the same word "marriage" for all these different packages of rights of obligations due to some sort of vague family resemblance between them.

A more sophisticated theory??refs may combine aspects of the social and individual theories, holding that not only do couples undertake obligations to each other and gain rights with respect to each other, they also undertake obligations to society and gain rights with respect to society.

But the individual and social theories are unsatisfactory for multiple reasons.

2.3.1. *Discovery.* People in good marriages come to discover new normative aspects to marriage as they go through life together. ??add-specifics? If the norms of marriage were simply whatever it was that the parties to the marriage chose, there would be nothing to discover. And if the norms of marriage were simply set by society, it would be odd to think that it is particularly by living the married life that one discovers the norms. Rather, the norms would be discovered by study of the history of the social institution of marriage, the laws surrounding it, the intentions of the legislators, and so on.

We might, admittedly, in individual and social institution cases discover new normative facts by logical derivation from previously known ones, but that is not actually the primary way in which we learn about marriage: we learn about it by observing it from the inside. And we discover new facts, including normative ones, about natural kinds of

entities precisely by observing these entities. By observing water, we come to see that it is H_2O and by observing mammals, we come to see that their middle ear should have the malleus, incus and stapes bones??check. And it is in our own case that we are best positioned to observe marriage at work, so it is unsurprising that such observation produces knowledge of normative aspects of the relationship.

Central to this growth is the Aristotelian harmony between different norms. Living according to the norms of marriage tends to fulfill us in other respects, while living contrary to the norms of marriage tends to be bad for us in other respects, and these are things we can often see. A happy marriage makes for happy spouses and an unhappy marriage for unhappy spouses.

2.3.2. *Travel.* Generally speaking, when a married couple emigrates to or visits another society, they are deemed married in their new place, unless there is some general reason that precludes them from counting as married, such as when they are of the same sex and move to a jurisdiction that does not recognize same-sex marriage.

Moreover, this recognition of them as married is not just an honorific indexed to their country of origin. When the Queen of Denmark visited the United States in 1991, she was referred to as a "queen"??check, but obviously she did not have rights and obligations of a monarch with respect to the United States, and so "queen" here was indexed with respect to the Kingdom of Denmark, and similarly for the title "prince" held by her husband. However, if someone referred to Henrik as Margrethe's *husband* or to Margreth as a *married woman* during the visit, these terms would not be merely indexed to Denmark. Rather, they would have the rights and obligations of an American married couple, as modified by their special immunity to persecution, and an ordinary non-diplomatic visitor from Denmark would not even have that modification.

Should we say that by the mere fact of entering a country, a couple that was married in their country of origin enters into a new marriage institution? On the social theory that is exactly what happens: the couple receives a new package of rights and obligations, definitive of marriage in a new society. But this would be quite surprising: it would mean that

a couple going for a honeymoon in another country would have had two weddings (one might tongue-in-cheek wonder if theyn they shouldn't then be entitled to a second honeymoon?), and globetrotting couples would rack up marriage after marriage. Moreover, relinquishing one's citizenship in a country one no longer lives in would be tantamount to a divorce.

Or perhaps instead of the new institution being entered into upon entry into a country, a couple by marrying enters into the marriage institution of every jurisdiction that is willing to recognize them as married, but the rights and obligations of these institutions are merely conditional on their being in those countries. While this would alleviate the problem of multiple weddings, such automatic entry into institutions in states that have no jurisdiction over one seems implausible. Moreover, the problem of multiple weddings is not solved. When Margrethe and Henrik married in 1967, there was no state of East Timor. Then in 1975 it declared independence. By that declaration, did they impose a new marriage institution on Margrethe and Henrik, a marriage institution that disappeared next year when East Timor was annexed by Indonesia, and then reappeared in 2002?

On the purely individual theory, the travel problem disappears. Different states may add rights and obligations, but what defines the marriage is the complex of rights and obligations that the couple entered into on their own, and it counts as a "marriage" in their travel destination because of the family resemblance between these rights and obligations and those that members of that society take on when they enter into an analogous relationship.

2.3.3. *Cross-cultural criticism.* Andronia is an especially sexist society, and Bob and Alice is an Andronian married couple. You've never interacted with Bob in a context that made his sexist views clear, but one day you find out that Alice is sick, and Bob is not showing any consideration for Alice besides the minimum needed to be shown to any human being. You call Bob out on this, and he tells you that in Andronia it is the wife's job always to show consideration for her husband while the husband need only keep the wife alive and show her the kind of consideration one owes every human being when she is

sick. He adds: "This is how my parents behaved, how Alice's parents behaved, and Alice knew that this is what she was getting into when we got married."

If marriage were a natural relationship, we could say that Bob and the rest of Andronian society is just wrong about what marriage requires, and we could say to Bob: "That may all be, but it's not how husbands *should* behave!" We could then show Bob examples of virtuous, caring egalitarian couples in the hope that these examples would open his mind to what marriage really entails. Or we might say to Bob: "If that's all you've committed to, then you're not really married, and so you are reaping the benefits of marriage from Alice under false pretenses."

But if the complex of obligations in marriage is either socially or individually defined, and if neither Andronian society nor the couple included any special obligation of husband to wife in sickness beyond that which we owe any other human being, then Bob could well be simply right in his understanding of his marital duties. This is an unattractive position.

Granted, if Bob and Alice are now living in a less sexist society, we could tell Bob that by moving to this society they have accepted the additional duties of husbands to wives. This is, however, dubious. It may be that be immigrating to a country we take on the legal obligations of that country, but it could well be the case that Bob is meeting these legal obligations, as they tend to be fairly minimal. What Bob is failing to do is to meet the customary obligations attached to marriage in less sexist societies, but it is implausible that by moving to a country one becomes obligated by the customs of the country. No moral criticism would necessarily attach to an American couple if after moving to Canada they failed to celebrate Thanksgiving in October. Furthermore, nothing of significance is changed in the above story if we specify that Bob and Alice are *still* living in Andronia. Be they in Andronia or elsewhere, a husband owes more to his wife than Bob thinks.

Perhaps we could tell Bob: "If that's all you committed to, then you aren't married in *our* sense of the word." But that isn't a criticism of Bob's behavior with regard to Alice. Bob could just say: "So what?" At most it is a criticism of his misuse of words if Bob claimed to be married. Moreover, even as a linguistic criticism it is unlikely to hold water. For we

do in fact use "marriage" and related words for relationships in a vast array of historical and present societies, many of which are quite sexist indeed.

Admittedly, on both the social and the individual view, we could criticize Bob for the relationship that he is in. We could tell him: "If that's what marriage in your context is, it's a corrupt institution, and you shouldn't be married to Alice." But this is unacceptably weak tea. It allows that Bob is married to Alice but does not owe her consideration in sickness beyond that owed a stranger.

2.3.4. *Fulfillment of a natural desire.* Plausibly, apart from reasonable moral and practical restrictions, people should be able to marry those whom they wish to. A society that did not make this possible would be failing its members.

Now, society has no obligation to make possible the fulfillment of every desire people have. Rather, it is reasonable to make a distinction between natural desires and more contingent desires, and hold that society should support the fulfillment of natural desires, such as for food, drink, shelter, useful employment, and knoweldge. Given the plausibility that marriage is one of those things society ought to make available to its members, it is plausible that the desire for marriage is a natural human desire. But if it is a natural human desire, then it is plausible that marriage itself is natural rather than constructed.

This is perhaps the weakest of the arguments for marriage being a natural relationship, however. First, not everyone shares the intuition that a society ought to make marriage possible. Second, it is not clear that we couldn't have a natural desire to construct—individually or socially—an institution of a certain type.

2.3.5. *Same-sex marriage.*

2.3.5.1. An argument for liberals. Let us assume that egalitarian justice requires one to advocate for same-sex marriage in jurisdictions where same-sex marriage is not available.

But suppose that marriage is socially constructed, and that we are in a locality in which one of the norms of marriage is that it be a relationship between a man and a woman. Then, if we understand "marriage" as the word is locally understood, it makes no more sense to

advocate for same-sex marriage than to advocate for chess without pawns: these are simply contradictions in terms. Granted, we may choose to advocate for social recognition of another, more egalitarian institution than marriage. But that will be a different institution.

If we advocate for this different institution, we have two choices. Either, we propose to maintain the institution currently called marriage, whether for everyone who wishes to enter into it or just for those grandfathered into it, or not. If we propose to maintain the current non-egalitarian institution, then we are not really advocating for same-sex marriage. We are advocating for a two-institution model, closely akin to marriage plus civil unions compromises that have generally been seen as unacceptable by advocates of same-sex marriage.

On the other hand, if it is proposed not to maintain the current institution of marriage, then the common and plausible arguments that extending marriage to same-sex couples does no harm to currently married opposite-sex will ring hollow. For it is a part of the proposal that the institution they are a part of be annihilated. Furthermore, in practice, in jurisdictions where marriage has been extended to same-sex couples, generally those who were previously married still count as married. Therefore, on the assumption that marriage is socially constructed, not only is there annihilation of the institution that couples used to be a part of, but these couples are, without their express consent, inducted into the new institution. Such automatic induction into a new relationship does not seem consistent with the ideals of a free society, and yet generally defenders of same-sex marriage have not been bothered by this.

If, however, one holds that marriage is a natural human relationship, then one can argue for marriage equality without arguing for a two-institution model or for the annihilation of the existing institution. Instead, one can hold that marriage is a natural human relationship which non-defectively can be instantiated by couples of the same sex as well as couples of the opposite sex. Given that marriage, understood univocally, can be entered in by both same-sex and opposite-sex couples, it is clear why it is discriminatory for a state to limit recognition of it to opposite-sex couples. And in advocating the end of

this inequality, one isn't advocating for an end of an existing institution, but simply for the state's recognition of the fact that this natural institution can equally well include same-sex and opposite-sex couples.

It is worth noting that defenders of marriage equality who hold that marriage is constructed by individual couples can also avoid the above problematic consequences of social construction. On the individual construction view, in recognizing a marriage, the state is recognizing is a certain type of contract, where the type is defined by a kind of family resemblance. But recognition of opposite-sex contracts of a certain type without recognition of same-sex contracts of a relevantly similar type would be unreasonable. Imagine if one could only sell a house to someone of the opposite sex, after all. On the individual construction view, the claim that no harm is done to opposite-sex couples by state recognition of same-sex marriage is easily defensible. So, the above argument from same-sex marriage advocacy supports the natural relationship view and the individual construction view, but not the social construction view.

2.3.5.2. An argument for conservatives. Here is a plausible principle: If we limit access to an institution on the grounds of gender or sex, absent very strong reason we should strive to make an equivalent available.??ref For instance, perhaps there is some reason for colleges to limit certain sports to one gender, but then they should make other sports available to the other gender. But many conservatives have not only object to same-sex marriage but also to the availability of civil-union institutions for same-sex couples. I will argue that such conservatives should embrace a view of marriage as a natural relationship.

For if marriage is constructed, either individually or socially, then even if the norms of that construction limit marriage to persons of the opposite sex, an equivalent institution without that limitation could be constructed, and by the principle at the top of this argument, it ought to be. In fact, it seems that the best way to resist this argument would be for the conservative to hold that marriage is a natural relationship, and that this relationship is only possible or only normatively possible for opposite-sex couples, while any

superficially similar relationship between persons of the same sex is not a natural relationship. Because no merely social institution would be a natural relationship, it would not be an equivalent to marriage. Therefore, the conservative can respond to the original argument by saying that there is very strong reason not strive to make an equivalent available, namely that no equivalent is possible.

In response, as per our previous argument, the defender of same-sex marriage should say that marriage is a natural relationship that *can* legitimately hold between persons of the same sex. So this conservative response does not close the debate. But it provides the conservative with a way forward. Indeed, it seems that both sides on the same-sex marriage debate will be better served by moving to a natural relationship view of marriage, and then discussing whether this natural relationship has norms that make it possible and permissible for persons of the same sex to instantiate it.

3. Double Effect

... intentions

... proportionality

... partiality

4. The task of medicine

The realism about teleology and normalcy provided by the Aristotelian framework allows for an elegant solution to the problem of what the task of medicine is.??ref:Lennox

The medical professional is a *professional*. Of course, everyone should refuse to act immorally on behalf of a client. But a professional has norms and pursues goals that go beyond general morality, and has reason to refuse to further the client's aims even when there is nothing generally immoral about these aims but the aims nonetheless violate the professional goals. Thus, while it is not immoral to create kitsch, a professional artist nonetheless has to refuse a commission that would be unavoidably kitschy.

In the case of some professions, the goals are very much socially defined, and apart from legal minutiae, the delineation of these goals is of relatively small importance. For instance, we have at least three professions that deal with the directing of water: gutter installers, sewer maintainers and plumbers. All three professions are important, but the division of labor between them is not of great importance. It would do little harm to society if we had a single profession for all three tasks, or if we divided up the tasks in some other way, say in terms of dealing with potable and non-potable water, or incoming and outgoing water relative to a house.

However, the division of labor between the medical professions and other professions does seem to cut nature at its joints. The medical professions directly aim at the goods of bodily health, a very natural subdivision of the space of human goods.

Moreover, there is a special value in the medical professional having a very sharp focus on health. Medical considerations are of great importance to everyone's life. But in the end, the patient (or their representative; I will simplify by talking just of patients) needs to be able to make a prudent decision about the recommendations from a medical professional, weighing this recommendation against non-medical considerations such as ones of economics, interpersonal relationships, personal pleasure and convenience, and so on. The patient is typically not an expert in biological matters, but tends to have a good grasp of other relevant goods: for instance, they will know what effect giving up alcohol would have on their social life, or what goods their children would have to give up if a medical procedure is to be paid for. It is important, however, that a medical recommendation be primarily concerned with the good of the patient's health, so that the weighing between medical goals and other goods be delegated to the patient as much as possible, and that the non-medical goods not be double-counted (once by the medical professional and again by the patient) in figuring out the prudent course of action.

At the same time, it is also important for guiding patients to prudent decisions that medical professionals understand health holistically, rather than narrowly thinking only of the kidneys or the feet. Thus, a focus on health in general is important for the medical

professional, or at least the medical professional who has an advising relationship with a patient.

But what is health? Health is not *simply* the good of the body. There are many goods of the body besides health, such as athletic prowess, beauty, and reproduction. These goods depend on health, but are not a part of health: for instance, a relatively healthy reproductive system is needed for reproduction, but one can have such a system without using it.

Apparently, physicians see their task as the return of the body to normal function, and then further claim to understand normalcy in a statistical way, as average function.??ref Tying health to normal function seems quite plausible indeed. But the normal cannot be understood merely statistically. ??xref? If it were merely statistical, then the adult who can deadlift 400 kilograms would be as abnormal as the adult who cannot deadlift one kilogram. Rather, normalcy often has a directionality that it inherits from a teleology towards some good. Adults who can deadlift 400 kilograms exceeds the norm, and might be said to be supernormal, but are not thereby abnormal, nor do they need medical treatment to reduce their strength.??Vonnegut

Moreover, among our goods, it is perhaps health that is most clearly species-relative. As a result, an Aristotelian metaphysics of forms is perfectly fitted to grounding the norms of health as the norms of sufficient capacity to function bodily in accordance with our human teleology.????more, better definition

5. Consent

?? http://alexanderpruss.blogspot.com/2022/08/a-tale-of-two-membranes.html

6. Environmental ethics

7. Relationship to other animals

8. 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:

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

(19) 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 (18), the oak tree naturally strives for reproduction. This natural activity of striving for reproduction, by (19), is then itself a perfection of the oak tree. Therefore, by (18), 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 (18). 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

[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.??

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 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

[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.

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 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

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

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. Balancing doxastic desiderata

I observe one raven, and it's black. I observe another and it's black, too. The story goes on. Every raven I observe is black. After a certain number of ravens, in a sufficiently broad number of settings, it becomes reasonable to believe that all ravens are black. But when?[1]

William James famously identified two incommensurable doxastic desiderata: attainment of truth and avoidance of falsehood. The larger the number of black ravens that are needed for me to believe that all ravens are black, the more surely I avoid falsehood, but the more slowly I attain truth. Intuitively, there is room for differences between reasonable people: some tend to jump to conclusions more quickly, while others are more apt to suspend judgment. But on either extreme, eventually we reach unreasonableness. Both someone who concludes that all ravens are black based on one observation and someone who continues to suspend judgment after a million broadly spread observations are unreasonable.

There is, thus, a range of reasonable levels of evidence for an inductive belief. And, as in the myriad of ethical cases of Chapter II, this raises the Mersenne question: What grounds facts about the minimum amount of evidence required for an inductive inference and the maximum amount at which suspending judgment is still rational? Of course, the "minimum" and "maximum" may may depend on the subject matter, on higher-order evidence such as about how well previous inductive generalizations have fared, and even on pragmatic factors(??ref). But that added complexity does nothing to make the Mersenne

[1] I am grateful to Sherif Girgis for raising the issue of incommensurable desiderata in connection with these issues.

question easier to answer. And, as we discussed in ??backref, invoking vagueness does not solve the problem, but multiplies the complexity even further.

And, of course, my contention will be that conformity to the human form is what grounds the answers for us. The rational way to reason is the way conforms to our form's specification of the proper functioning of our intellect.

It appears to be quite plausible that different answers to the rationality questions would be appropriate for species of rational animals adapted to different environments. First, some possible worlds as a whole have laws of nature implying a greater uniformity than that found in other worlds, and hence make it appropriate to make inductive inferences more quickly. Second, the environments that the rational animals evolved in may have greater or lesser uniformity, despite the same laws of nature. Third, the ecological niche occupied by the rational animals may punish falsehood more or may reward truth more. ??explain with examples Because of this, the Aristotelian species-relative answer to the Mersenne questions is particularly appealing.

2. Logics of induction

Attempts have been made to give precise answers to the questions about the reasonableness of inductive inferences using a rigorously formulated logics of induction.??refs Let us suppose, first, that some such logic, call it L_{12}, does indeed embody the correct answers. Nonetheless, we will have a Mersenne question as to why L_{12}, rather than one of the many alternatives, is the logic by which we ought to reason inductively.

In the truthfunctional deductive case, there is a system that appears to be both particularly natural and matches our intuitions so well that it has gained a nearly universal following among philosophers, logicians, mathematicians and computer scientists: two-valued boolean logic. It is a sociological fact that no logic of induction has anything like this following, and a plausible explanation of this sociological fact is that no logic of induction has the kind of naturalness and fit with intuition that would privilege it over the

others to a degree where it would seem non-arbitrary to say that it is *the* logic we should reason with.

Further, observe that logics of induction can be divided into two categories: those with parameters (say, parameters controlling the speed of inductive inference–??refs) and those without.

A logic of induction with parameters raises immediate Mersenne problems about what grounds the fact about which parameters, or ranges of parameters, are in fact rationally correct.

A parameter-free logic of induction, however, is not likely to do justice to the fact that different ways of balancing rational goods are appropriate in different epistemic and pragmatic contexts. Moreover, it is unlikely to do justice to the intuition that the balancing should be different in different species of rational agents.

3. Goodman's new riddle of induction

All the emeralds we've observed are green, and it's reasonable to infer that all emeralds are green. But Goodman's famous riddle notes that all the emeralds we've observed are also grue, but it's not reasonable infer that all emeralds are grue. Here, an emerald is grue if it is observed before the year 2100 and green, or if it is blue and unobserved. According to Goodman, the predicate "is green" is *projectible*, i.e., amenable to inductive inference, while the predicate "is grue" is not. But how do the two differ?

As Goodman notes, the fact that "grue" is defined in terms of "green" and "blue" does not help answer the question. For if we specify that something is bleen if it is observed before 2100 and blue, or it is never observed and blue, then we can define something to be green provided it is observed before 2100 and grue or never observed and yet bleen, and similarly for "blue" with "grue" and "bleen" swapped.

Whatever the *justification* may be, it is clear that induction with "green" is reasonable, but not so with "grue". Notwithstanding Goodman's symmetry observations, "grue" is

a gerrymandered predicate, as can be seen in accounting for it in terms of more fundamental physical vocabulary. But now observe that "green" is also gerrymandered. An object is green provided that the wavelength profile of its reflected, transmitted and/or emitted light is predominantly concentrated somewhere around 500 to 570 nm. The actual boundaries of that region are messy and appear vague, the measure of predominant concentration is difficult to specify, and accounting for reflective, transmittive and emissive spectra is a challenge. The full account in terms of more fundamental scientific terms will be complex and rather messy, though not as badly as in the case of "grue", which is more than twice as complex since it needs to account for blueness and the rather messy date of "2100", which is quite a messy date in more fundamental physics units (perhaps Planck times since the beginning of the universe?). Where the boundary between non-projectible and projectible lies—what counts as too gerrymandered for projectibility—is an excellent Mersenne question.

There is a very plausible way to measure the degree of gerrymandering of a predicate. We take a language the content of whose symbols are terms for fundamental physical concepts, or more generally concepts corresponding to fundamental joints in reality, and we look for the shortest possible formula logically equivalent to the predicate, and say that the predicate is gerrymandered in proportion to the length of this formula. It is indeed likely that by that measure "is grue" is more than twice as complex "is green".??ref:Lewis

But now notice something odd. Say something is "pogatively charged" if it is positively charged and observed before 5×10^{60} Planck times or never observed and negatively charged. All the protons we have seen are pogatively charged. But we should not conclude that all protons are pogatively charged. It seems that "is pogatively charged" is just as unprojectible as "is grue". However, notice that by the formula length account, "is green" is more gerrymandered than "is pogatively charged". Pogative charge is much closer to the fundamental than colors. It seems, thus, that our Mersenne question about the boundary between the non-projectible and projectible is not merely defined by a single

number—a threshold such that predicates definable with a length below that number are projectible.

Perhaps, however, what is going on here is this. The hypothesis that all emeralds are grue cannot overcome the hypothesis that all emeralds are green, even though both fit with observation. Similarly, the hypothesis that all protons are pogatively charged cannot overcome the hypothesis that all protons are positively. So perhaps rather than an absolute concept of projectibility, we have a relation of relative projectibility: "is green" is projectible relative to "is grue" and "is grue" is non-projectible relative to "is green".

We can once again try to account for this in terms of the complexity of formulae. But now we need to compare the complexity of two formulae. And where previously we had a single numerical threshold as our parameter of projectibility, we now have a threshold and a new non-numerical parameter that specifies the mathematical way in which the complexities of the two terms are to be compared. This parameter specifies how we test against the threshold: the ratio of complexities, the difference in complexities, or some other mathematical function of the two complexities?

Furthermore, while the idea of a language all of whose terms reflect fundamental joints in reality can be defended, the grammar of the language will make a difference to the precise complexity measurements. For instance, if we have the fundamental predicates Cx, Dx and Ex, then the complex formula expressing the predicate "is C as well as either D or E" will be

$$Cx \,\&\, (Dx \vee Ex)$$

in infix notation, and hence five times longer than the formula Cx represening "is C", but in Polish notation will be

$$KCxADxEx$$

and hence only four times longer than Cx.

For a relative projectibility relation defined in terms of linguistic complexity, we thus have at least three free parameters, each a fit subject for a Mersenne question: a threshold, a comparison function, and a grammar for the basic language.

But in fact we probably should not think of a binary projectible / non-projectible distinction, whether relational or absolute. As Goodman himself observed??ref-in-https://www.jstor.org/stable/pdf/686416.pdf, what we have instead is a range of predicates that are more or less projectible. We have "is green" and "is grue". But we can also say that x is grue* provided that x is green and observed by a French speaker before 2100 or by a non-speaker of French before 2107, or not observed, and "grue*" will be less projectible than "is grue". On the basis of our observations, the probability that all emeralds is green is very high, and the probability that they are all grue or grue* is very low. But nonetheless, the probability that they are grue is somewhat higher than that they are grue*. After all, an alien conspiracy to recolor emeralds upon observation with a sharp cut-off in one year seems a little bit less unlikely than one where the cut-off depends on whether the observer speaks French. Similarly, it makes sense to think of "is green" as less projectible than "is positively charged", and of "is cute" as even less projectible.

Projectibility now becomes a matter of degree. An advantage of this is that perhaps we no longer need to make it relational. The reason for the superiority of the green-hypothesis to the grue-hypothesis and for the positive-charge-hypothesis to the pogative-charge-hypothesis can be given in terms of the relationship between the degrees of projectibility. However, the cost is that now we need a function from predicates to degrees of projectibility, and the choice of that function will have infinitely many degrees of freedom.

4. Epistemic value

4.1. Epistemic value on its own. Plausibly, the more sure you are of a truth, the better off epistemically you are, and similarly the more sure you are of a falsehood, the worse off you are.

But what exactly is the dependence of value on the degree of certainty? Fix some hypothesis H and let $T(p)$ be the epistemic value of having degree of belief or credence p (where $0 \leq p \leq 1$) in H if H is in fact true and let $F(p)$ be the value of credence p in H if H is in fact false. The pair T and F is called an accuracy scoring rule in the literature.??ref

We can put some plausible constraints on T and F. First, $T(p)$ cannot decrease if p increases, and $F(p)$ cannot increase if p decreases.[2] But that still leaves infinitely many degrees of freedom for the selection of T and F.

We can, however, make some progress if we reflect on expected values. If your current credence in H is p, then by your lights there is a probability p of your having epistemic score $T(p)$ and a probability $1 - p$ of your epistemic score being $F(p)$, so your expected score is:

$$pT(p) + (1 - p)F(p).$$

Suppose now you consider doing something odd: without any evidence, brainwashing yourself to switch your credence from p to some other value p'. By your current lights, the expected epistemic value of this switch is:

$$pT(p') + (1 - p)F(p').$$

And this shouldn't be higher that the expected epistemic value of your actual credence p. For surely by the lights of your assignment of p to H, no other credence assignment should be expected to do better. Indeed, if another credence assignment p' were expected to do better by the lights of p, then p would be some kind of a "cursed probability", one such that if you assign it to H, then immediately expected value reasoning pushes you to replace it with p'. This is not rational. So, it is very plausible indeed that:

$$pT(p) + (1 - p)F(p) \geq pT(p') + (1 - p)F(p').$$

[2] We might more strongly specify that $T(p)$ always strictly increases with p, and $T(p)$ strictly decreases. That is plausible, but one might also have a view on which there is a finite number of discrete thresholds at which increase/decrease happens.

If T and F satisfy this inequality for all p and p', we say that the pair T and F is a *proper* scoring rule. And if by the lights of the assignment of p to H, that assignment has better expectation than any other, i.e., if the inequality above is strict whenever $p \neq p'$, we say that the rule is *strictly proper*.

Propriety reduces the degrees of freedom in the choice of scoring rule. Given any non-decreasing function T, there is a function F that is unique up to an additive constant such that the parir T and F is a proper scoring rule, and conversely given any non-increasing function F, there is a T unique up to an additive constant such that T and F is a proper scoring rule.??? Hence, once we have one of the two functions, the other is almost determined. However, at the same time, this result shows what a profusion of proper scoring rules there is: for every non-decreasing function, there is a proper scoring rule that has that as its T component.

The question of epistemic value assignment may seem purely theoretical. However, it has real-world ramifications. Suppose a scientist has attained a credence p in a hypothesis H, and is considering which of two experiments to perform. One experiment will very likely have a minor but real effect on the credence in H (think here of a case where you've gathered 1000 data points, and you now have a chance of gathering 100 more). The other will most likely be turn out to be irrelevant to H, but there is a small chance that it will nearly conclusively establish H or its negation. For each experiment, the scientist can use their present credence assignments to estimate the probabilities of the various epistemic outcomes, and can then estimate expected epistemic values of the outcomes.

It is well-known??ref that if the scoring rule is strictly proper, for each experiment that has potential relevance to H (i.e., there is at least one outcome that has non-zero probability by the scientist's current lights and learning which would affect the credence in H), the expected epistemic value of performing the experiment is higher than the expected epistemic value of the *status quo*. Thus if the experiments are cost-free, it is always worth performing more experiments, as long as we agree that the appropriate scoring rule is strictly proper, and it does not matter which strictly proper scoring rule we choose. But if in addition to

deciding whether to perform another experiment, the decision to be made is *which* experiment to perform, then the choice of scoring rule will indeed be important, with different strictly proper scoring rules yielding different decisions.??ref:fill-in

There are a number of mathematically elegant strictly proper scoring rules, such as the Brier quadratic score, the spherical score and the logarithmic score. Of these, the logarithmic score is the only that is a serious candidate for being *the* correct scoring rule, in the light of information-theoretic and other arguments (??ref:phil of sci paper). In our setting where we are evaluating the value of a credence in a single proposition H, the logarithmic score is $T(r) = \log r$ and $F(r) = \log(1 - r)$.

However, there are also reasons to doubt that the logarithmic score is the One True Score. First, there is an immediate intuitive problem. If you are certain of a falsehood, your logarithmic score is $\log 0 = -\infty$, while if you are certain of a truth, your score is $\log 1 = 0$. Now, while there is good reason to think that the disvalue of being sure of a falsehood exceeds the value of being sure of a truth, it is somewhat implausible that it infinitely exceeds it.

For the next two problems, note that logarithmic scores and the arguments for them only really come into their own when we are dealing with more than two propositions (in our above setting, we had H and $\sim H$ are the only relevant possibilities). Suppose we are dealing with n primitive possibilities or "cells", $\omega_1, \ldots, \omega_n$ (say, the sides of an n-sided die), and that our agent has assigned credence p_i to ω_i. If in fact ω_i eventuates, the logarithmic score yields epistemic value $\log p_i$.

One of the merits touted for the logarithmic score is that ???for how many cells??? (up to multiplicative and additive constants) it is the only proper score where the epistemic value depends only on the credence assigned to the cell that eventuates. But this is also a serious demerit. Suppose that you and I are trying to figure out how many jelly beans there are in a jar. Let's say that our range of possibilities is between 1 and 1000. I look very quickly and assign equal probability $1/1000$ to each number. You count very carefully and arrive at 390. But then you think that although you are really good at counting, you might

be off by one. So you assign 998/1000 to 390, and 1/1000 to each of 389 and 391. It turns out that the number is 391. We both have the same logarithmic score, $\log(1/1000)$, since we both assigned the same probability $1/1000$ to cell 391. But intuitively your assignment is much better than mine: you are better off epistemically than I.

Finally, observe that in real life, credences are not consistent—do not satisfy the axioms of probability. And the logarithmic score allows one to have extremely inconsistent credences and still do well. If I assign credence 1 to *every* possible outcome, I am guaranteed to max out the logarithmic score no matter what. Thus one of the least rational credence assignments results in the best possible score.

We now have two different approaches to the Mersenne questions about epistemic value and scoring rules. First, we could suppose that there is such a thing as *the* One True Score. Since only the logarithmic score seems significantly mathematically privileged over all the other scores, and the logarithmic score is not the One True Score, there will be an appearance of contingency about the One True Score even if there is one.

Second, we might suppose that just as rational people can differ in prudential preferences, they can differ in epistemic preferences. Some may, for instance, have a strong sharpish preference for gaining near-certainty in truths, while being fairly indifferent whether their credence in a truth is 0.6 or 0.8, as neither is that close to certainty. Others, on the other hand, may value increased certainty in a gradual way, like the logarithmic rule does.

However, it is important to note that while there may be room for rational people to differ in epistemic preferences, there is reason to think that there are rational constraints on epistemic preferences that go beyond formal conditions such as strict propriety, continuity or symmetry—where the last is the condition that $T(p) = F(1-p)$.

Let $T_0(x) = 1000$ if $x \geq 0.999$, $T_0(x) = -1000000$ if $x \leq 0.001$, and $T_0(x) = 0$ otherwise. Let $F_0(x) = T_0(1-x)$. Then the pair T_0 and F_0 is a symmetric and proper scoring rule.??ref

Consider now a scientist who adopts this scoring rule for some hypothesis H of minor importance about some chemicals in her lab that she initially assigns credence $1/2$ to. She

has a choice between two methods. She can use clunky machine A that she has in her lab, which is guaranteed to give an answer to the question of whether H is true, but for either answer there is a 0.11% chance that the answer is wrong. Or she can use spiffy new machine B which has the slightly lower 0.09% chance of error either way. The only problem is that her lab doesn't own machine B and her grant can't offer the price. Her only hope for using machine B is to go and buy a scratch-off lottery ticket which has a one in a million chance of yielding a prize exactly sufficient to purchase machine B. However, because some chemicals involved in the experiment are expiring exactly in a week, and machine A is slower than machine B and takes exactly a week to run, if she is to use machine A, she needs to start right now and doesn't have time to buy the lottery ticket. And once she starts up machine A, she can't transfer the experiment to machine B.

In other words, her choice is between using machine A, and then learning whether H is true with a credence of 0.9989, or buying a lottery ticket, which gives her a one in a million chance of learning whether H is true with a credence of 0.9991 and a 999,999 out of a million chance of being no further ahead. Going for the second option seems irrational if all that is at stake is epistemic value: the difference between 0.9989 and 0.9991 is just not worth the fact that most likely going with the lottery route one won't learn anything about H. (If what was at stake wasn't epistemic value but something pragmatic, then things could be different. We could imagine a law where some life-saving medication can be administered to a patient only if we have 0.9991 confidence that it'll work, and then there will be no practical difference between $1/2$ and 0.9989, but a big one between 0.9989 and 0.9991.)

But a scoring rule like the one described above prefers the lottery option. For the epistemic value of using machine A is guaranteed to be zero since after using machine A, the scientist will have credence 0.9989 or 0.0011, depending on whether the result favors H or not.

On the lottery option, however, conditionally on winning the lottery, the expected epistemic value will be:

$$(1/2)(0.9991 \cdot T(0.9991) + 0.0009 \cdot F(0.9991)) + (1/2)(0.9991 \cdot F(0.0009) + 0.0009 \cdot T(0.0009)) =$$

since it is equally likely given the scientist's priors that the machine will return a verdict for or against H, which will result in a credence of 0.9991 or 0.0009, respectively, and in either case there will be a 0.0009 chance that the verdict is erroneous. Since $F(0.0009) = T(0.9991) = 1000$ and $T(0.0009) = F(0.9991) = -1000000$, it follows that the expected epistemic value, conditionally on winning the lottery, will be:

$$(1/2)(0.9991 \cdot 1000 + 0.0009 \cdot (-1000000) + 0.9991 \cdot 1000 + 0.0009 \cdot (-1000000)) = 99.1 > 0.$$

And if we multiply this by the 1/1000000 chance of winning the lottery, we still have something positive, so the expected epistemic value of playing the lottery with the plan of using machine B is positive, while that of using machine A is zero.

Thus, by considerations of epistemic value, the scientist with this scoring rule will prefer a 1/1000000 chance of gaining credence 0.9991 as to whether H is true to a certainty of gaining the slightly lower credence 0.9989. This is not rational.

Now, in the above example, our scoring rule while proper, symmetric and finite, was neither continuous nor strictly proper. However, we will show in the Appendix??ref that there is a sequence of continuous, strictly proper, finite and symmetric scoring rules T_n and F_n such that $T(x) = \lim_{n\to\infty} T_n(x)$ and $F(x) = \lim_{n\to\infty} F_n(x)$ for all x. If n is large enough, then the pair T_n and F_n will require exactly the same decision from our scientist as T and F did, since the expected value of the expected (T_n, F_n)-scores of the two courses of action will converge to the expected vlue of the expected (T, F)-scores.

Hence not all epistemic valuations that satisfy the plausible formal axioms are rationally acceptable, then we will have Mersenne questions about what grounds the further

constraints on the epistemic valuations. These constraints are likely to include messy prohibitions, with multiple degrees of freedom, on the kinds of sharp jumps that our pathological scoring rule above exhibited.

Furthermore, things become more complicated when we consider that the epistemic value of a credence in a truth will differ depending on the importance of that truth. Getting right whether mathematical entities exist or whether we are material or how life on earth started has much more epistemic value than getting right Napoleon's shoe size. Epistemic value will thus not only be a function of credence and truth, but also of subject matter. Moreover, we will have further degrees of freedom concerning the operation of combining epistemic values for different propositions—addition may seem a plausible operation, but the logarithmic and spherical rules are not combined additively across propositions.

We thus have multiple indicators of a contingency about epistemic value assignments. And there is good reason to think that different forms of life are more suited to different epistemic value assignments. The most obvious aspect of this is that once we move away from the toy case of assigning a value to one's epistemic attitude to a single proposition and consider that attitudes to a large number of propositions need to be considered, it is obvious that the subject matter of the propositions will affect how great a weight we give credences about them in the overall evaluation. And the importance of subject matter obviously depends on the form of life. It is plausible that for intelligent agents whose natural environment is more hostile it would be more fitting to have a greater epistemic value assigned to practical matters, while agents that have few natural enemies and can get food easily might more fittingly have a greater epistemic value assigned to theoretical matters. One imagines here that intelligent antelope might be properly expected to be less philosophical than intelligent elephants.

4.2. Connection with other values. ????????????

5. Bayesianism

5.1. Introduction. Bayesianism is the best developed picture of what a precise and rigorous account of epistemic rationality would be like. It is thus worth looking carefully at what kind of answers the Bayesian could give to the questions we have been asking.

??introduce Bayesianism

5.2. Priors. From a Bayesian point of view, how induction works is determined by the probabilities prior to all evidence, the ur-priors. Suppose, for instance, that I assign equal prior probability to every logically possible color sequence of observed ravens. For simplicity, suppose that there are only two colors, white and black. I find out that there are a million ravens, and I observe a thousand of them, and find them all black. I am about to observe another raven. The probability that the next raven will be black will be $1/2$. For the sequence $B, ..., B, W$ (with 1000 Bs) is just as likely as the sequence $B, ..., B, B$ (with 1001 Bs), and both sequences fit equally well with our observations.

On the other hand, suppose I assigned probability $1/3$ to the hypothesis that all ravens are white, $1/3$ to all black, and split the remaining $1/3$ equally among the $2^{1000000} - 2$ multicolor sequences. My observation of the first 1000 ravens then rules out the all-white hypothesis. And it rules out most of the multicolor sequences: there are $2^{999000} - 1$ multicolor sequences that start with 1000 black ravens, which is a tiny fraction of the original $2^{1000000} - 2$. Since as a good Bayesian I keep the ratios between the probabilities unchanged, each of the remaining multicolor sequences has $1/(2^{1000000} - 2)$ of the probability of the all-black sequence, and since there are only $2^{999000} - 1$ multicolor sequences remaining compatible with the evidence, the ratio between the multicolor probability and the all-black probability is $(2^{999000} - 1)/(2^{1000000} - 2)$ to 1, or approximately 1 to 2^{1000}. Thus, we have overwhelming confirmation of the all-black probability, and hence an even more overwhelming confirmation of the hypothesis that the next raven will be black.

Other ways of dividing the probabilities between the hypotheses yield other results. Carnap??ref, for instance, had a division that worked as follows. For each number n between zero and a million we have the hypothesis H_n that there are exactly n black ravens, and Carnap proposed that all million-and-one of these hypotheses should have equal probability, and then each hypothesis H_n is divided into equally likely subhypotheses specifying all the subhypotheses that make there be n ravens. Thus, H_0 and $H_{1000000}$ have only one subhypothesis: there is only one way to have no-black or all-black. But H_1 and H_{999999} have a million subhypotheses each: there are a million options for which is the raven with the outlying color. Using the same constant-ratio technique as before, after observing 1000 black ravens, the chance that the next one is black will turn out to be approximately 0.999, but the chance that all million are black will only be 0.001. More generally, if there are N ravens, and the first m of them have been observed to be black, and $n \geq m$, then the probability that the first n will be black will be $(1+m)/(1+n)$.[3] Hence we have very reason to think that the *next* raven is black, but unless we have observed the bulk of the ravens, we won't have reason to think that all the ravens are black.

Intuitively, while Carnapian probabilities support induction, they result in induction being too slow—it is only when we have observed the bulk of the cases being a certain way that we get to conclude that they are all like that. My 1/3–1/3–1/3 division is too fast. Even with 1000 black ravens having been observed, the probability of a white raven shouldn't be *astronomically* small in the way that $1/2^{1000}$ is. Reasonable priors, thus, yield a speed of induction somewhere between these.

We can presumably come up with a formula for the priors which will fit with our intuitions of how fast induction should work. For instance, we could take Carnap's setup, but increase the prior probability of the all-white and all-black raven hypotheses. But such an increase would be apt to involve one or more parameters. If the specific assignment of

[3]Let B_m be the claim that the first m ravens are black. Then $P(B_m) = \sum_{n=0}^{N-m} \binom{N-m}{n} / ((N+1)\binom{N}{m+n}) = \frac{1}{1+m}$.??why:Mathematica The probability that the first n are black given that the first m are black where $n \geq m$ will then $P(B_n \mid B_m) = (1+m)/(1+n)$.

priors were rationally required of us, then we would have the Mersenne question of why it is these and not some other very similar priors that are required. And if there is a range of priors rationally permitted to us, then we would have Mersenne questions about the boundaries of this range.

Further, imagine beings other than us that inhabit a more Carnapian world than we do. While in our world, we have a significant number of natural kinds that exhibit or fail to exhibit some basic property exceptionlessly—for instance, every electron is charged, and no photon has mass–in that world there are few such natural kinds. Instead, if we were to tabulate the frequencies of basic binary properties in various natural populations—say, tabulating the frequency of blackness among ravens, charge among electrons, mass among photons—we would find the frequencies to be distributed uniformly between 0 and 1. In that world, Carnapian priors would lead to the truth faster than the more induction-friendly priors that we have. And let us imagine that in that world we have intelligent beings who reason according to Carnapian priors. Even if we happily grant that Carnapian priors are irrational for us, it seems plausible to think that they could be rational for those beings. To insist that these Carnapians are irrational, because it would be irrational for us to have these priors, seems akin to saying that bigamy would immoral for aliens who need three individuals to reproduce, or that there is something wrong with fish because they lack lungs.

Consideration of the rationality of induction thus once again reveals an appearance of contingency in the normative realm, which once again yield an argument for an Aristotelian picture of human nature, where the rationally required priors or ranges of priors are those that we are impelled to by our human nature.

But before we embrace this conclusion fully, we should consider two Bayesian challenges, from two opposed points of view. The algorithmic Bayesian thinks that considerations of coding can yield a reasonable set of priors, while the the subjective Bayesian says that there are no constraints on the priors.

5.3. Algorithmic priors. Suppose, first, we have a language L of finite sequences of symbols chosen from some finite alphabet of basic symbols, with some of the sequences representing a member of some set S of situations.

For instance, S could be arrangements of chess pieces on a board[4], and L could be a declarative first-order language with no quantifiers, twelve piece predicates (specifying both the color and piece type) and sixty-four names of squares. We could then say that a symbol sequence represents an arrangement a provided that the sequence is a syntactically valid sentence that is true of a and of no other arrangement. However, in general L need not be a declarative language. It could, for instance, be an imperative computer language for an abstract Turing machine or a physical computer, and the situations could be possible outputs of that machine. Then we might say that a symbol sequence s represents a possible output a just in case s is a program that, when run, halts with the output being a. Or if we like we might add an additional layer of representation between the outputs of the machine and the situations—for instance, the outputs of an abstract Turing machine might represent the physical arrangement of particles in a universe. We can even chain languages. For instance, we could have a computer language L_1, with the outputs being sequences of symbols in some declarative language L_2, whose sentences in turn represent members of a set S of situations.

Next, consider a natural way of choosing at random a finite sequence of symbols of L. Here is one. Add to L's finite alphabet a new "end" symbol. Then randomly and independently, with each symbol being equally likely (i.e., having probability $1/(n+1)$, where n is the number of non-end symbols), choose a sequence of symbols until you hit the end symbol. The sequence preceding the "end" symbol will then count as the randomly selected sequence in L. Every sequence of length k has probability $1/(n+1)^k$, so the probabilities decrease exponentially with the length of the sqeuence. We repeat the random selection

[4]The arrangements contain less information than chess positions, since a chess position includes other information, such as whether a given king or rook has already moved, whose turn it is, as well as historical information needed for adjudicating draws.

process until we get a sequence that is both syntactically correct and represents a situation in S.[5] We now stipulate that the prior probability of a situation s is equal to the probability that the above process will generate a sequence that represents s.

Alternately, we can formulate this as follows. If a is a sequence of symbols of L and s is a situation in S, write $R(a,s)$ if a is syntactically correct and represents s, and let $R(a)$ be shorthand for the claim that a syntactically correctly represents S, i.e., $\exists s(s \in a \ \& \ R(a,s))$. Then if A is a randomly chosen sequence of symbols of L, we can define the prior probability $Q(s)$ of s as the conditional probability that A syntactically correct represents s on the supposition it syntactically correctly represents something, i.e.,

$$Q(s) = P(R(A,s) \mid R(A)) = \frac{P(R(A,s))}{P(R(A))}.$$

We can call these L-Solomonoff priors.

These priors favor situations that can be can be more briefly represented in L over ones whose representations are long. The effect of these priors depends heavily on the choice of language L and how well it can compress some situations over others. For instance, in our chess case, if we have no quantifiers, it is easy to see that any two piece arrangements with the same number of pieces will have equal prior probability, because each square's contents have to be separately specified. Thus, if we know that every square contains a pawn, and we have observed the first 63 of these pawns and found them all to be black, the probability that the 64th square will be black is still $1/2$. On the other hand, if quantifiers and identity are allowed into our language, then the all-black-pawn situation can be briefly represented by

(1) $\forall x(\text{BlackPawn}(x))$

(where the domain is squares on the board), while the situation where squares $1, ..., 63$ have black pawns and square 64 has a white pawn is harder to represent. We might, for instance, use a sentence like:

[5] If at least one finite sequence is syntactically correct and represents a situation in S, then with probability one, we will eventually get to a sequence that syntactically correctly represents some sequence.

(2) $\forall x(\sim(x = 64) \rightarrow \text{BlackPawn}(x))$ & $\text{WhitePawn}(x)$.

Since the probability of generating a given sequence of symbols decreases exponentially with the number of symbols,

(3) i

s much less likely to be randomly generated than

(4) ,

and it is intuitively very likely (though proving this rigorously would be quite difficult???refs) that in general the probability of generating a sentence representing 64 black pawns is higher than that of generating a sentence representing 63 black pawns followed by one white pawn. We can thus expect that the conditional probability of the 64th pawn being black on the first 63 being black to be very high. (Maybe even too high? It is very difficult to get good estimates here, because there are many ways that a single situation can be represented. ??refs)

Just as in the case of using linguistic complexity to quantify projectibility, the choice of language here provides many Mersenne questions. If we opt for the algorithmic version of the theory, we need to choose some computer language for a real or abstract computer, and then we need to choose a representation map between outputs and situations in the external world, with infinitely many possible candidates. And on the more descriptive versions, we still need to choose a language, with many decision points as to syntax and vocabulary. It is very unlikely that there is a privileged language. And not every will fit with our intuitions about induction. For instance, we can easily create a language, whether algorithmic or descriptive, where 63 squares of black pawns followed by one square with a white pawn are much more briefly describable than 64 squares with black pawns. For instance, on the descriptive side, we might use a $\text{BlitePawn}(x)$ predicate, where something is a blite pawn provided it is a black pawn and on one of the first 63 squares or a white pawn and on the 64th, and an analogous $\text{WhackPawn}(x)$.

In the unlikely case that there is a privileged language L such that L-Solomonoff priors are rationally required for us, we will have a vast number of Mersenne questions about the various parameters of the language and its representation relation. In the more plausible case that there is a set of languages such that we are required to have L-Solomonoff priors for some L in the set, we will have a vast number of Mersenne questions about the parameters that control the range of languages. All of this gives rise to a significant degree of appearance of contingency.

Consider, too, the following observation. It is implausible that the languages defining rational priors for us should be ones that are completely beyond our ken. But, on the other hand, it is plausible that there are possible languages that to the smartest human are as incomprehensible as one of the less intuitive computer languages like Haskell or Verilog or one of the creations of logicians further from natural language like lambda-calculus is to a typical six-year-old. Imagine now beings to whom such these languages beyond human ken are easy. It *could* be the case that the norms for rational priors for them are formulated in terms of L-Solomonoff priors for one of the "baby languages" that humans can understand, but this does not seem a particularly plausible thesis. It seems more likely that for those beings, the algorithmic rational priors would be different than for us.

The standard way??ref to defend algorithmic measures of complexity from the problems presented by a plurality of languages is to observe that sufficiently sophisticated languages have translational resources. Thus, one can write a Haskell interpreter in Javascript, and so anything that can be expressed in Haskell can be expressed in Javascript by including the code for a Haskell interpreter, and then using a string constant that contains the Haskell code. The result is that the difference in the length of code needed to generate a given output in different computer language will typically not be more than an additive constant: if one can produce the output with Haskell code in n bytes, then one can produce it in approximately[6] $n + k_{H,J}$ bytes in Javascript, where $k_{H,J}$ is the length

[6]The approximation is due to complications due to having to embed code in a string constant, which may involve various escape characters.

of the Haskell interpreter in Javascript. For large enough n, the additive constant will be unimportant. If we are to measure the complexity of a one-hour broadcast-quality video by the length of code needed to compress the video, the addition of $k_{H,J}$ will likely be negligible: a Haskell interpreter is about half a megabyte, while an hour of video compressed losslessly can be reasonably expected to be in the gigabytes.

However, two points need to be made in our epistemological context. First, even if two different languages give very similar sets of priors, if they give even slightly different priors, we either have the Mersenne question of what makes one of these sets of priors be the objectively correct one, or we have the Mersenne question about the boundaries of the range of permissible languages. Second, unlike in the case where we are measuring the complexity of a large set of data, such as a video file, in the inductive cases we need to look at ways of expressing relatively simple statements, such as "All electrons are negatively charged" or Schrödinger's equation or "The first 63 squares have a black pawn and the last square has a white pawn." But for such statements, a translation manual will dwarf the length of the translated text. To say "All electrons are negatively charged" in French by first describing how English works, and then saying that this description should be applied to the English sentence, will produce a French sentence that is many orders of magnitude longer than the English one, and hence not a sentence that is relevant to measuring the prior probability that all electrons are negatively charged.

Finally, while there is something elegant and natural about randomly choosing items in L by randomly choosing within the set of symbols with an end marker added, there are other ways to proceed. For instance, instead of making the end marker equally likely as each of the ordinary symbols, one could at each step of generation flip a fair coin. On heads, one is done generating. On tails, one then uniformly randomly chooses one of the n symbols.[7] Or one might first randomly choose a positive integer specifying the length

[7]This will actually not make a difference in those languages where the syntax already determines where a syntactically valid sequence ends. This will be the case with some Polish notation languages, where a valid sequence ends when the main operator is filled out with arguments.

of the sequence of symbols according to some probability distribution on the positive integers, and then make all the sequences of that specified length be equally likely. Or one might randomly choose a positive integer, and then choose the nth sequence of symbols in some ordering (e.g., alphabetical). While the initial symbol-by-symbol method with an end-symbol may seem more elegant, it is hard to say that it is rationally privileged to the point that the priors generated with it are rationally required. But if it's not thus privileged, then the range of random choice methods will provide more Mersenne questions. For not every random choice method yields priors that are plausible candidates for rational permissibility. There will be random choice methods where the sentence "Birds are a government-run drones" is many orders of magnitude more likely than all other sentences taken together, and so a boundary would need to be posited between the admissible and inadmissible random choice methods.

On a final note, ne might think that the intuitively most natural way of choosing a linguistic item at random is to make them all be equally likely. Unfortunately, this presents serious mathematical and philosophical difficulties. For a language based on finite sequences taken from a finite (or countable) alphabet, there are countably infinitely many sentences: we can enumerate them $s_1, s_2, s_3, ...$ in some arbitrary way. But if each one is equally likely, with probability some real number α, then we have a problem. In classical probability, we will have to have:

$$1 = P(\{s_1\}) + P(\{s_2\}) + P(\{s_3\}) \cdots = \alpha + \alpha + \alpha + \ldots.$$

But if $\alpha > 0$, then the right-hand-side is infinite, while if $\alpha = 0$, it is zero, and in neither case is it 1. There are technical ways of escaping this by departing from classical probability. They all require restricting the additivity axiom of probability that says that if $A_1, A_2, \ldots$ are countably many disjoint events then the probability of the union of the events is equal to the sum of the probabilities of the events to the case where there are only finitely many events. After that, one either takes $\alpha = 0$ or takes α to be a positive

infinitesimal—something that is bigger than zero and smaller than any positive real number.

But whatever one does on the technical side, there will be philosophical difficulties. Emblematic of them is this paradox. Suppose you and I play a game where we each randomly pick a sentence with all sentences equally likely, and without seeing the other's sentence. When I see my sentence, whatever it turns out to be, I inevitably become nearly sure that your sentence comes after mine in the sequence, because there are only finitely many sentences that come before mine and infinitely many that come after. And you come to be convinced of the same thing. This leads to paradoxical decision-theoretic conclusions. For instance if we are playing a game where one wins if one has a sentence further down in the sequence, you will then be willing to pay me any amount short of the prize to swap sentences with me, no matter what sentence you got. ?????

In any case, the equal probability approach itself does not appear to be a good way to generate induction-friendly priors, because it lacks the preference for shorter descriptions that is central to the functioning of algorithmic or linguistic priors.

Once we abandon, as we should, the equal probability approach to choosing a linguistic item, anything else is a matter of choosing from among infinitely many ways to be biased in favor of shorter expressions. Some of these are mathematically more elegant than others, but none is decisively so.

5.4. Subjective Bayesianism. Subjective Bayesians avoid all the difficulties of specifying permissible priors by merely requiring the priors to satisfy some formal properties. These are taken to include the axioms of probability and, sometimes, the regularity constraint that all contingent propositions have non-zero probability. Rationality then constrains transitions from one set of probabilities to another: these must follow the Bayesian update rule that upon receiving evidence E, one's probability in a hypothesis H goes from $P(H)$ to $P(H \mid E)$. But the initial choice of priors is up to the individual, subject to the formal constraints.

The resulting picture of rationality does not match common sense. Take the most ridiculous set of conspiracy theories that we would all agree is unsupported by our evidence, but where nonetheless the conjunction of the theories is logically consistent with the evidence. Then there is a possible assignment of priors such that updating in the Bayesian way on our actual evidence strongly confirms the conjunction of these theories, both in the incremental sense of greatly increasing that probability and in the absolute sense of making that probability high. (All we need is that the prior probability of the conjunction of theories be low, but the conditional probability of that conjunction on the conjunction of our evidence be high.) Yet to reason that our evidence strongly supports these theories is paradigmatic of irrationality. It shouldn't be the case that a rational person could come to exactly the same conclusions on exactly the same evidence as are paradigmatic of irrationality.

Or consider the implausible asymmetry between the freedom in choosing initial priors and the rigid constraint in updating. Suppose I don't like my current set C of posteriors, and I would feel better if I had some cheerier alternative set A of posteriors. There is some set of priors Q that, given the evidence E that I had received over my lifetime, would have yielded A. According to the subjective Bayesian there would have been nothing irrational in having adopted those priors in the first place, and thus having ended up at the cheerier posteriors. Why should I be tied to the priors that I actually had?

The picture of priors here is like a rationally unbreakable vow to live one's life as an evolution of these priors under Bayesian application of evidence. But it is a vow made without any rational ground, indeed without any choice, and likely in childhood. We would think it unsupportable that someone be held committed for life to a promise made early in childhood. Why then should we be bound to our priors and the posteriors coming from them?

5.5. Non-Bayesian update. Bayesianism is committed to updating by conditionalization being the only permissible way to update credences. If my current credence in a hypothesis H is $P(H)$, and I receive evidence E, my credence should move to $P(H \mid E) =$

$P(H \ \& \ E)/P(E)$. No other changes of credence are permitted. There is an elegance and non-arbitrariness to this, of course modulo the above-discussed issues about the choice of the priors $P(H)$ and $P(H \mid E)$.

But as is the case for many other formally simple philosophical theories, this is too simple. Consider several cases.

PILL: A trustworthy oracle offers you a pill which will shift your credences closer to truth and does not introduce any incoherence.

MISTAKE: An hour ago, I made an arithmetical evidence when updating on evidence. I moved from credence 0.4 in H to credence 0.6, even though in fact $P(H \ \& \ E)/P(E)$ equalled 0.7. I have just discovered my mistake. Surely it would be good for me to go back and correct my credences, essentially rewinding my epistemic life from the mistake and re-conditionalizing on all the subsequent evidence.

STUPIDITY: My original priors had extremely high probabilities for some conspiracy theory involving members of a certain minority group, so high that despite the fact that all my life I have been receiving strong evidence against the theory, nonetheless I was quite convinced of the theory. I reflect on my original priors, and note that my priors of the conspiracy theory in question were quite out of proportion to my priors for similar theories involving other minority groups. I conclude that my original priors were stupid and racist. I go back and change them to treat the various groups more equally, and then as best I can I fix up my posteriors to match the evidence I recall having had.

In all three cases, my update of credences seems quite rational, but is not a case of conditionalization. We should thus suppose that while it may be a good general rule that we should update by conditionalization, we should at times depart from it. But the question of how we should depart from conditionalization now introduces complexity in the theory that raises significant Mersenne questions. We need more than just simple exceptions for each case: there will be parameters to set.

The pill case as given is straightforward. We can say that you should depart from Bayesian update when doing so would move some of your credences closer to truth and

none further away from it, where a credence's "distance from truth" is the distance of the credence from 0 or 1 depending respectively on whether the proposition the credence is in is false or true. However, there are variants of the pill case. Suppose that the pill has a 0.9 chance of moving you significantly closer to truth and a 0.1 chance of moving you slightly away from truth. It seems like it could be worth it. Or suppose that the pill moves your credences in important propositions closer to truth at the expense of moving your credences in some unimportant propositions further from the truth.

This suggests that we will need to quantify epistemic value, much as we did in Section??backref, and then say something like this: You are epistemically required (respectively, permitted) to opt for a pragmatically costless change of credence when doing so increases (does not decrease) expected epistemic value. If we quantify epistemic value in terms of strictly proper scoring rules, a delightful result of this move is that we do not need to handle the ordinary case of Bayesian update any differently. For it can be proved that in the kinds of cases where one simply receives evidence and sets one's credences according to it, conditionalization is the unique best policy for maximizing expected epistemic value.??ref:(Also Isaacs and Russell??)

However, as we also saw in Section??backref, the choice of a scoring rule or measure of epistemic value involves infinitely many free parameters, and at the same time not every scoring rule that satisfies the formal constraints is rationally plausible. In our present context of evaluating non-conditionalizing updates the point is particularly clear. For when evaluating credence-changing pills, we need to take into account the *importance* of the affected credences, and that is not a formal criterion.

Note that pill-type cases are not as outlandish as they may initially seem. We do in fact modify people's thinking with psychiatric medication as well as with psychotherapy.

Next, consider the case of the mistake in updating. We don't in fact remember all our evidence: we update our credences on the more important things and forget the less important, which often includes some or all of the evidence. I know that Beijing is the capital of China, but I don't know where I learned it from. I know that there was a cat in

the yard earlier this morning, but much of the rich sensory evidence is now gone from my mind. An attempt to go back and correct a past update error is bound to be a messy affair. Sometimes when the mistake is minor it might be better to let it go rather than miss out on some of the evidence gathered since. And the situation is often too messy and too complex for any sort of an expected epistemic utility calculation.

Yet the fact that a situation is messy does not get us off the hook. There will still be a distinction between right and wrong ways to proceed, even if they cannot be formalized. We have here something very similar to the kinds of messy everyday ethics cases that are grist for the situationist's mill.??backref-or-add It is unlikely that the answer is given by any elegant principle without the kinds of parameters that raise Mersenne questions, but at the same time there is an answer, and there must be something to ground it.

And what goes for the messiness in correcting a calculational mistake applies even more to the messiness involved in the kind of intellectual conversion that occurs when one realizes that all of one's thinking for years has been based on irrational prejudice, and one attempts to dig oneself out of the resulting heap of epistemic defects. Again, there are right and wrong ways to proceed, but the likelihood of a clear and elegant principle that solves the problem is nearly nil.

There is a final set of issues with update. Presumably, there is something to the maxim that ought implies can. And typically we can't do Bayesian update. We don't have the time, don't have the mathematical skills, or perhaps most importantly aren't able to quantify our priors in such a way as to make them amenable to precise mathematics. We need an epistemology that works in this all too human predicament, a *human* epistemology, and we need an account of what grounds that epistemology.

One may insist that under these imperfect conditions, we should simply say that we are stuck with irrationality. But nonetheless there are cases where it is clear that something should be done under the unhappy circumstances. Suppose that the completely right credence on my priors in some proposition p is 0.9874, but I can only calculate to two significant figures. Then I should take the credence to be 0.99, not 0.98 or 0.01. But things

will be less clear if I need to form credences in both p and q, which a perfect Bayesian with my priors and evidence would take to be 0.9874 and 0.0332, but due to limitations of time or energy level, I can only get a total of four significant figures. Are the right credences for me 0.99 and 0.03, or 0.987 and 0.0, or 1 and 0.033? It depends, surely, on the relative epistemic importance of p and q. And once importance needs to be taken into account, it is very unlikely that we will have a precise and elegant account with no free parameters. Instead, we have human messiness.

6. Intellectual limitations

Limitations in recognizing logical implications and contradictions are a clear case where rational norms are at least species-relative. It is epistemically irrational for one to conclude that it's raining from the fact that it's neither raining nor snowing. But it is not irrational for one to take $22828 \times 2219 = 50645332$ (which is in fact false) to follow from the axioms of arithmetic due to an arithmetical slip. Of course, someone with exceptional calculational skills may immediately that see the latter is mistaken, but the ordinary person's failure to see it is not an instance of irrationality.

Ethics also contains intellectual limitation cases. Analogically to the multiplication case, we would not consider a person to be less than virtuous because they are unable to see an extremely complex medical ethics case rightly. But if someone doesn't realize that it's wrong to ambush strangers in order to sell their organs, there seems to be something morally wrong.

But the ethical cases may also be different. Consider an adult who strives to follow their conscience as best they can, but nonetheless fails to see that it's wrong to kill strangers for their organs. This person presumably has a severe intellectual and/or emotional disability. We might judge them to be a good person, or at least not a vicious one. However, the logical case seems different. The person who, due to intellectual disability, concludes that it's raining from the claim that it's neither raining nor snowing *is* failing at rationality, though of course they are inculpable. One can defensibly define moral worth in terms of

the seriousness of their attempt to do what seems right, but it is difficult to define someone's degree of rationality in terms of the seriousness of their attempt to think as seems right. For it is paradigmatic of irrational people that they take themselves to be acting quite rationally—if one derives that it's raining from its neither raining nor snowing, this is precisely because the conclusion seems to follow. It is, indeed, unclear whether it is even possible to conclude p from q without its seeming that p follows from q. But it is paradigmatic of immoral people that they lack integrity and violate their own conscience.

The line to be drawn in the epistemic rationality cases (and in the ethical ones if they turn out to be similar), is a line that we are unlikely to be able to draw in a species-independent way. It seems plausible that for beings that as a species are more adept at logic than we are, a failure to see the truth of Fermat's Last Theorem as following from the axioms of arithmetic is a failure of rationality, but it is not so for us.

Recently, Jeffrey??ref has argued that ethical norms may be relative to a stage in life. Whether this is so, it is very plausible that some epistemic norms are so. There may be logical facts failure to notice which constitutes a failure of an adult's rationality, but would not constitute a failure of a child's rationality. If this is right, then we have all the more reason to think the norms of epistemic rationality to be species-relative, since surely what the stages in life are, and when they occur, is something that is species-relative.

6.1. Innate beliefs and testimony. Humans are said to have much less in the way of instinct than other earth animals. Furthermore, we do not seem to have any innate beliefs. But we can imagine a species of intelligent animals which do more by instinct than we do, and which have evolved to be born with some unshakeable and true beliefs, such as that purple winged things are to be avoided and electrically charged spiky fruit is good to eat. For these beings, there is nothing irrational about having such beliefs without any evidence. For us, there is. In a Bayesian mode, we might say that for these beings very high priors in these empirical claims are appropriate, but not so for us.

We can also imagine a species where memories are inherited. A member of this species could, then, be born with a large number of beliefs that they had no evidence for. For, like

we often do, the members of this species could forget the evidence that led to a conclusion. But while in our case, we once had the evidence, in this species it was some ancestor of theirs that had the evidence. It is plausible that for us it is generally irrational to have beliefs without ever having had evidence. But this is just a normative fact about our species, not a fact about species-independent rationality.

Now, let us return to our species. Perhaps we should not think the species where memories are inherited to be all that different from us. For do not our parents bequeath much knowledge to us? It is true that this is mediated by vibrations of the air rather than by gametes, but does that make a significant difference. We should, thus, take seriously the idea that just as for members of a species where memories are inherited it is fundamentally rational to believe these inherited memories, and to question them without special reason is irrational, for us it may be fundamentally rational to believe at least some testimony, and to question it without special reason is irrational. If so, then we have another example where the Bayesian way of looking at update may not be correct. ???fill out

7. Going beyond the applicability of human epistemology

Suppose that you are certain you are one of infinitely many people have each rolled a fair die, none of whom have seen the result. It is then announced by a perfectly reliable angel that all but finitely many of the dice show six. What should be your credence that your die shows six?

On the one hand, it seems very likely that your die shows six. After all, the vast majority of the dice show six, and you have no reason to think yourself exceptional.

On the other hand, prior to the announcement, your credence in six was $1/6$. And the announcement told you nothing about your die. For the following two statements are logically equivalent:

(5) All but finitely many of the dice show six.

(6) All but finitely many of other people's dice show six.

For your die's state makes no difference to whether there are finitely or infinitely many non-sixes. But then given your certainty that the dice are fair, (6) gives no information about your die's result. And since (5) is logically equivalent, it too gives no information. Having received no information, you should stick to 1/6.

This reasoning seems convincing. Yet if everyone sticks to 1/6, then all but finitely many people are quite far from the truth. Furthermore, if everyone is playing a game where you are asked to guess if you have a six or not, and you are rewarded for getting it right and penalized for getting it wrong, then if everyone sticks to 1/6 for having a six, everyone will guess that they don't have six, and all but finitely many people will be penalized, which is surely not the right result.

In ??ref, I argued that a good solution to epistemological paradoxes like this is to reject the metaphysical possibility of an event depending causally on infinitely many events, in the way that the angel's accouncement depends on the infinite number of die rolls. But there is another possible solution: we can deny that our epistemic norms extend to far-fetched situations, just as in ??ref it was suggested that our ethical norms do not apply to far-fetched situations. Whether this is a completely satisfactory resolution to the paradox is not clear. For there is some plausibility in thinking that if predicaments like the above were metaphysically possible, there could be rational beings who could reason in them. But it seems there couldn't be any. However, no matter what we say about the metaphysical possibility of such beings, there is something right about the thought that *we* are not made to reason about such things.

There are many other examples where epistemology seems to break down in radical cases. What should you think if you came to be convinced that an evil demon is trying to get you to believe as many falsehoods as possible? Any thought you might have ends up undercut. Should your credences in everything be at 1/2, then? But that is incoherent: for then you have credence 1/2 that the last die you rolled is 1, and credence 1/2 that it was 2, and credence 1/2 that it was 3, and credence 1/2 that you never tossed a die, and so on. Perhaps suspension of judgment is something other than assigning probability 1/2.

But if so, should you suspend judgment about everything, including about the norm of fitting your beliefs to you evidence? However, if you suspend your judgment about that, then why bother with suspending your judgment about other things, given that you don't think you need to fit your beliefs to your evidence? Or what should you think if you come to be convinced that you are a computer simulated non-player character in a sophisticated video game? What kind of a world should you think you are in?

Note now that there is such a thing as realizing that when you were using certain words, you had no concept behind them, and the words were mere meaningless words. For instance, take a word that we as laypeople defer to experts on the meaning of, such as "gluon". But now imagine that physicists have been pulling our collective legs about gluons all this time: it was just a made-up word without any meaning behind it. We can imagine discovering that. And words whose meaning we get by deference are not the only words like that. The phenomenon of a person who doesn't know what they are talking about is not uncommon. My metaphysics includes accidents. But I could imagine finding myself in a position where I come to be convinced that I never had a concept of an accident—that "accident" is a meaningless word. Now imagine a radical hypothesis: I come to be convinced that I have no concepts at all. What should I think now?

There is something plausible about the thought that in certain extreme situations there is no right way for us to think. ??other beings*

8. Facts about species-independent rationality?

??:also, species-independent ethics

rationality-and-ethics constitutive stuff?? God??

9. Meta-epistemology

Appendix: *Approximating the pathological scoring rule with continuous ones

We need to show that the stepwise scoring rule (T, F) from ??backref can be written as a limit of symmetric, strictly proper, finite and continuous scoring rules.

First note that any symmetric continuous proper scoring rule (t, f) can be written as the limit of symmetric continuous strictly proper scoring rules by letting $t_n(x) = t(x) - (1-x)^2/n$ and $f_n(x) = f(x) - x^2/n$, since the Brier scoring rule defined by the functions $-(1-x)^2$ an $-x^2$ is strictly proper, and the sum of a proper and a strictly proper scoring rules is strictly proper.

Thus, all we need to show is that is that we can approximate (T, F) with symmetric, finite and continuous scoring rules. Furthermore, we can drop the symmetry requirement. For write $f^*(x) = f(1-x)$. Then (t, f) is a proper scoring rule if and only if (f^*, t^*) is a proper scoring rule. Now if $T(x) = \lim_n t_n(x)$ for all x and $F(x) = \lim_n f_n(x)$, then

$$(T(x) + F^*(x))/2 = \lim_n (t_n(x) + f_n^*(x))/2$$

and

$$(F(x) + T^*(x))/2 = \lim_n (f_n(x) + t_n^*(x))/2.$$

But $T = F^*$ and $F = T^*$, so the left-hand sides are just $T(x)$ and $F(x)$, respectively. Moreover, $(t_n + f_n^*, f_n + t_n^*)$ is will be a continuous symmetric finite proper scoring rule if (t_n, f_n) is a continuous finite proper scoring rule.

Fix $\varepsilon > 0$. Let ϕ_ε be a continuous non-negative finite function that is zero except on the set $U_\varepsilon = [0.999 - \varepsilon, 0.999) \cup (0.001, 0.001 + \varepsilon]$, and is such that $\int_{0.999-\varepsilon}^{0.999} (1-x)\phi_\varepsilon(x) = 1000$ and $\int_{0.001}^{0.001+\varepsilon} x\phi_\varepsilon(x) = 1000000$. Define

$$T_n(x) = \int_{1/2}^{x} (1-u) phi_{1/n}(u)\, du$$

24 and

$$F_n(x) = \int_{1/2}^{x} u\phi_{1/n}(u)\, du.$$

By ??SchervishThm4.2, (T_n, F_n) is proper. It is clearly continuous and finite. It is, further, easy to calculate that $T_n(x)$ and $F_n(x)$ equal $T(x)$ and $F(x)$ except perhaps on $U_{1/n}$.??check For any x in $[0, 1]$, there is an N such that x is not in $U_{1/n}$ for any $n \geq N$. It follows that $T(x) = T_n(x)$ and $F(x) = F_n(x)$ if $n \geq N$, and we have the limiting condition we wanted.

CHAPTER VI

Mind

1. Naturalistic options

1.1. Multiple realization.

1.2. Functionalism, malfunction and evolution.

2. Teleology and representation

3. Teleology and mental causation

4. Teleological animalism

4.1. Animalism.

4.2. Cerebra.

5. Soul and body ethics

CHAPTER VII

Semantics

1. Communication and norms

1.1. A problem about cooperation. ??cut:squeaking is better?

There are scenarios, such as the Prisoner's Dilemma or the Tragedy of the Commons??Refs, where it is difficult to see how to rationally secure cooperation between agents. The following should not be one of these. You have two agents who will each in a separate booth choose whether to press a red button or a blue button. If they both press the same button, they each get a reward, say a chocolate bar. If they press different buttons, they each get a penalty, say a nasty electric shock, with the penalty outweighing the award by a significant factor, so it's better to get neither than to get both. If either player omits to press a button, neither gets anything, and the buttons are so set up that one cannot press both. Moreover, the players are allowed to confer ahead of time.

Obviously, when conferring ahead of time, they will need to decide which button to press, by rolling a fair die or flipping a fair coin if necessary, and then they need to go into their booths and press that button. Neither has any incentive to defect to pressing the other button, and there is no risk in pressing a button should the other defect fail to press anything. This is a really easy win-win game.

But now suppose our two players, Alice and Bob, are perfect expected utility maximizers who break ties with fair coinflips, and the only relevant utilities are the rewards and penalties of the game. There are no further games that will be played. Nobody outside the game is in any way affected by the results (e.g., nobody will be disappointed if one of them breaks a promise). And because each player gets the same payoff, it won't matter whether Alice and Bob maximize collective utility or their own personal utility. Finally, the above

information is completely luminous to both players. I claim that at this point the obvious strategy—to decide on a button and then both press it—is no longer rationally available.

For concreteness, let's suppose that Alice and Bob have agreed to press the red button. They go into their booths. What will Alice do? She is a perfect expected utility maximizer. She will only press the red button if the expected utility of doing so is at least as big as that of all the alternatives (these being being pressing the blue button or pressing no button). Now the expected utility of pressing the red button is only going to be at least as big as the expected utility of pressing neither button if Alice takes it to be significantly more likely that Bob will press red the button than that Bob will press the blue button.[1]

But why should Alice take it to be significantly more likes that Bob presses the red button than the blue button? Ordinary human beings take themselves to be beholden to norms of promise-keeping, and tend to abide by those norms, especially when there is no obvious benefit to failing to do so. But Bob is a pure expected utility maximizer. Whatever normative force he takes promises to have has to be derivable from the norm of expected utility maximization. In ordinary contexts, dealing with ordinary human beings, keeping promises certainly does maximize utility, because ordinary human beings believe in norms of promise-keeping and punish those who break those norms (if only by castigating or refusing to enter on joint projects with promise-breakers). But Bob is not dealing with an ordinary human being. He is dealing with an expected utility maximizer.

Here is one way to see the difficulty. Imagine that Alice and Bob are perfect utility maximizers with a perverse value theory that in addition to common-sensical value assignments to chocolate bars and electric shocks assigns non-instrumental negative value to keeping promises and non-instrumental positive value to doing the very opposite of what one has promised. I will stipulate that pressing the button of the other color counts

[1]Suppose Alice maximizes only her own utility (if she maximizes collective utility, just double all the utilities). Suppose $x > 0$ is the reward and $-y < 0$ is the penalty with y bigger than x by a significant factor. Then the expected utility in pressing the red button will be $\alpha x - \beta y$. Alice will only press the button if $\alpha x - \beta y \geq 0$, i.e., if $\alpha/\beta \geq y/x$. Since y is bigger than x by a significant factor, this requires α to be bigger than β by a significant factor.

as the "very opposite". In this case, if there is joint knowledge of the perverse value theory, it is more reasonable to expect Alice and Bob to press the button other than the one they promised. And now imagine that they are perfect utility maximizers with a value theory that assigns positive value to keeping promises and non-instrumental negative value to doing the opposite. In this case, it will be more reasonable to expect Alice and Bob to press the button they promised. But then the in-between case, where Alice and Bob are perfect utility maximizers and assign zero value to promise-keeping and promise-breaking, should be one where the probabilities of pressing the promised button and pressing the opposite button are equal.

What if we suppose that the solution here is that as a matter of contingent fact people have a preference for promise-keeping over promise-breaking: we feel bad when we break promises and good when we have fulfilled them. Preferences enter into utilities, and so if Alice and Bob have the standard preferences, they will have a bias in favor of promise-keeping, and if each knows the other to have the preference, then each can take the other's preference into account, and hence each can expect the other to keep the promise.

First, it is not clear if this solves the problem if we imagine the penalty for mismatched button presses increased so that a preference for avoidig the penalty is an order of magnitude stronger than the preference for promise-keeping. In that case, unless Alice and Bob are going to be very confident in the other's choice, a preference for promise-keeping will not do the job.

Second, and more importantly, if a preference for promise-keeping is needed to solve the problem, we now have an argument that some norm encumbent on humans requires such a preference. For if two human beings are stuck in a suboptimal solution in the button-pressing game, they are clearly falling short of what humans should be able to achieve. The argument thus shows that there must be norms on human beings that go beyond utility maximization, whether collective or individual.

1.2. Arresting the regress of meaning. Some communicative actions—speech acts or gestures—have their significance assigned through earlier communicative actions. Thus,

sometimes one coins a word and stipulates its meaning in terms of other words, sometimes one uses gestures to introduce a new word, and sometimes one just hopes that use in a rich enough communicative context will clarify the meaning. But barring outlandish hypotheses such as that humans got their language from aliens, who got theirs from an infinite regress of angels, we cannot suppose an infinite regress. There must be ur-communicative actions, ones which did not get their significance from earlier communicative actions.[2]

At the same time, there is a contigency here. While it feels natural to us to use an extended index finger to indicate the nearest salient object approximately along the ray extending from the knuckle in the direction of the fingertip, it would be possible to have rational beings that use this gesture to indicate the third-nearest salient object along the ray extending from the index finger's tip to the knuckle. There is no necessary connection between the physical behavior and its significance. We thus have a Mersenne question here: What explains the correlation between physical behavior and significance in ur-communicative actions?

We might try to explain the correlation in terms of the actual contingent behavior of individuals and communities which arises by natural or social selection. Suppose, for instance, that some social animals evolve to squeak at a certain pitch when observing a predator, thereby warning other members of their group. Eventually, their descendants develop rationality, but the squeaking behavior is maintained, and remains correlated with the presence of a predator, even though it is now under voluntary and rational control. It is plausible to say that the squeaking is now a communicative activity whose significance is "Predator!"

[2]There is a Sellarsian objection to this. Perhaps there are behaviors prior to the advent of rationality that count as having communicative significance in virtue of *later* behaviors, in a kind of virtuous significance-conferring circle.??ref If so, however, then we can just count the whole circle of behaviors as the first ur-communicative community action.

But this plausible claim deserves more careful examination. Rationality complicates things. Suppose Alice sees a predator and is considering to squeak to trigger her groupmates' defensive behavior. If Alice's squeaking and her fellows' defensive behavior is to be rationally chosen, the agents need reasons. The fact that their ancestors used to squeak when predators were present and start defensive behavior upon such squeaking is an interesting bit of pre-history, but there does not appear to be a reason for them to imitate this quaint custom. Even if we add the fact that they find themselves with a desire to squeak in the presence of the predator and to initiate defenses upon hearing a squeak, these desires at most generate the very weak kinds of reasons one has to fulfill miscellaneous subrational desires rather than the strong kinds of reasons that one has to warn one's fellows and protect oneself and one's cmmunity.

There is no difficulty here if Alice is the only rational one, and hence the others will act on instinct. Alice can then rationally squeak to trigger the instinct. Similarly, if Alice acts on instinct and the others are rational, they can infer from her instinctive behavior that there is a predator present, as one infers fire from smoke. But when both are acting purely rationally, we have a difficulty. Likewise, if Alice thinks there is a fairly high chance that her fellows will follow their habit and prepare themselves, or if she knows that her fellows think there is a fairly high chance that Alice would find herself squeaking in the throes of instinct, there is no difficulty. The difficulty shows up when we have nothing but rationality at play.

Perhaps we can solve the problem by positing a non-rational preference for squeaking when seeing a predator and for preparing a defense when hearing a squeak? After all, arguably, even a perfectly rational being will act in accordance with preferences when other things are equal.

But the non-rational preference is insufficient here unless it is implausibly strong. For even if one finds oneself with an urge to squeak in the presence of a predator, the squeak itself endangers one. Because of this, for a rational being, the brute preference for squeaking is not sufficient to motivate the squeak. It is only when one thinks the squeaking will

trigger defensive behavior among one's fellows that it's worth squeaking. Similarly, we may suppose that defensive behavior is costly and inconvenient, and is only worth engaging in, not withstanding the non-rational preference, when there is reason to think there is an actual predator.

Instead of a brute preference, perhaps convenient priors will do. Thus, suppose that members of the community simply find themselves with a high prior conditional probability that a member of the community rationally squeaks when presented with a predator and does not squeak when not presented with a predator.[3] Knowing about that this prior is wide-spread in the community, Alice can squeak in order to get her fellows to update their credences in favor of a predator. The priors seem pleasantly self-confirming: if community members have the priors, then it will become public that they do so, and the squeaking behavior will match the priors.

But now suppose Bob is reflecting on his convictions. Bob finds himself accepting a correlation between Alice's rational squeaks and the presence of predators. But since the squeaks are rational, there must be a rational explanation of these squeaks. Since we are no longer attempting a preference-based story, presumably the explanation is that Alice accepts a correlation between community members hearing squeaking and their rationally coming to think there is a predator. In other words, Bob finds himself having a brute prior concerning a contingent and empirical matter—namely, another community member having a certain credential state. However, when we find out that our conviction about a contingent and empirical matter is simply a brute prior, that tends to undermine the conviction. Suppose that I find myself believing there is vast treasure buried under my house. I search for the source of my belief, and find it's just a prior. Absent a story such as that angels put that prior in my head to encourage me to dig out the treasure, finding out that this was *just* a prior should undermine the confidence, contrary to what subjective Bayesians think.

---

[3] I am grateful to ?? for the suggestion that priors might do the job.

Couldn't natural selection play the angel, though? There is an adaptive advantage to correlated priors in Alice and Bob that make communication possible, and these priors then end up automatically matching reality. ???

1.3. Reason-generating mechanisms. The transition from non-rational signalling to rational communication is thus difficult to analyze. We need some sort of a reason-generation mechanism.

Can we suppose that the reason-generation mechanism here is a necessary one? Perhaps it is a necessary truth that when there is a pre-rational behavior that tends to be triggered by circumstances C, then that behavior when done rationally *signifies* C? But there would be multiple Mersenne questions that would be raised by such a necessary truth. First, we need to select one item C in the chain of causes rather than another—does the squeak signify the predator's, or the light in the air between the predator and the observer's eyes, or the immediate cause of the predator's presence? There are multiple selection rules, no one of them significantly more natural than the others. Second, what reliability does the tendency have to have in order to yield a signification fact? ??more The parameters in the connection between behavior and significance point to something contingent. We can imagine different species of rational beings where the parameters are different from what they are in us.

Reasons are normative entities. Thus a contingent reason-generation mechanism will, plausibly, be a mechanism for generating norms. Aristotelian form fits well here. The form could directly specify that squeaking properly occurs only when there is a predator present, or it could specify a general rule for connecting pre-rational behavior with norms of significance.

But even if there is such a norm-generating process, what makes the norms be norms of *communication*? A cat's nature requires it to turn its ears towards relevant sounds. When we see a cat turn its ears in some direction, that provides us with evidence that there was some sound relevant to it. But the cat is not communicating that there is a sound relevant to it by turning its ears. What, then, makes it be the case that a norm in the nature of

a communicative animal is a communication-constituting norm? Do we not need some further primitives besides norms of proper function to make it be *communicative* proper function?

We can speculatively sketch a part of an answer in terms of the *content* of norms. A toy story could be that some norms come in pairs, where one norm posits that a certain overt behavior is only proper when some fact p is known to a community member to obtain and another community member is known to be present and capable of observing the behavior, and another norm posits that when that overt behavior is observed in another member of the community, there is a tendency to form a belief in p. In that case, the toy story says that a behavior that is a fulfillment of the first norm counts as a communication of p and a behavior that is a fulfillment of the second norm counts as a reception of p. Of course, the full story would need to be much more complicated.

And all that said, it is not clear that we need a full story as to which exact behaviors are in fact communications. What matters for figuring out what to do is the content and force of the norms, not what kind of norms it is. It is a tautology that if the force of a norm is kept fixed, the norm has the same reason-giving impact on us, whether it be a norm of semantics, prudence, etiquette or morality.

2. Content and indetermincy of reference

Wittgenstein, Kripke, Quine and Putnam??refs have problematized reference and content in light of the fact that different content attributions to our locutions can be made to fit with our behavior. The Wittgenstein-Kripke line of thought notes that any finite number of cases of behavior can be made to fit with infinitely many rules. Any finite number of utterances of "$a + b = c$" that fit with our "usual" interpretation of "$+$" will also fit with infinitely many rules, including, say, the rule that "$a + b = c$" means that c is identical with a plus b when a and b are less than or equal to x and means that c is identical with a times b when at least one of a and b is bigger than x, where x is the largest number we have ever discussed in the context of "$+$". The Quinean line of thought observes that the

same word, "Gavagai", can be interpreted to mean a rabbit or an undetached rabbit part, with both interpretations fitting equally well with the community's practices. And finally the Putnamian line of thought observes that a remapping of the truth conditions can make "The cat is on the mat" mean any other true proposition, as well as noting that the identities of mathematical objects—such as the integers—would be underdetermined even by a countably infinite number of statements about them.??ref ??expand-and-exposit

In all of these cases, we have the initial intuition that there is a well-defined meaning to the locutions, an intuition that is destabilized by the arguments. These cases, thus, can be seen as the opposite of the cases of vague terms like "bald", where our initial intuition is that there is no well-defined meaning.

We thus have two families of arguments. One family of arguments pushes in the direction of indeterminacy. And it does so not just in the cases where indeterminacy is intuitive, as for "bald" and "heap", but alas also in cases where we expect determinacy, as with the question whether we are referring to rabbits or undetached rabbit parts. Another family of arguments, mainly those based on insistence on classical logic??backref, push in favor of determinacy, but alas also in cases where we expected indeterminacy. If we want to maintain the determinacy of pretty much any term, then we will need to hold to something somewhat problematic—we will need to bite the bullet by denying a premise of a relevant indeterminacy argument (plausibly, some variant of the Quinean argument applies to all terms). But similary if we want to maintain the indeterminacy of any term, we will have to wrestle with classical logic.

With regard to these arguments, it would be simpler either to embrace determinacy in all cases or to embrace indeterminacy in all cases. For then we would only need to bite the bullet on one set of arguments. If we are to do this, then embracing determinacy in all cases seems preferable—embracing indeterminacy about all of language seems like it could undercut too much of our practices. However, by treating all the cases alike, we go against common sense which distinguishes "bald" from "rabbit".

The Aristotelian has a particularly good hope of having a metaphysical answer to the arguments for indeterminacy. This can be embraced in all cases, thereby resulting a picture of a sharp world that we will discuss below, or only in some cases, which fits with common sense.

The arguments for indeterminacy are all based on an assumption that the correct semantic theory will make semantic facts supervene on facts about our actual behavior and the world around us. But we can reject this assumption, and add normative facts about humans to the facts about our actual behavior and the world around us as part of what the semantic facts supervene on. These further facts could be hyperintensional normative facts, such as that it is only appropriate to say "Gavagai!" in the presence of a rabbit. Granted, necessarily, one is in the presence of a rabbit if and only if one is in the presence of an undetached rabbit part. But there can still be a difference between the norm of its being appropriate to say "Gavagai!" in the presence of a rabbit and a norm of its being appropriate to say it in the presence of an undetached rabbit part.

For norms are hyperintensional: ϕing and ψing might be such that necessarily one does one if and only if one does the other, but it is still a different thing to be required to ϕ than to be required to ψ.??Faroldi One way to see this is that if one is required to do something, one is required to try to do it. But trying, like intending and believing, is clearly hyperintensional. It is a different thing to try to bisect or trisect an angle with ruler and compass than to try to bisect an angle with ruler and compass, even though, necessarily, one bisects or trisects if and only if one bisects, since trisection is impossible. If I do not know that trisection is impossible and I have promised a friend to show them a trisection or bisection, what I am obligated to try is different than had I promised to demonstrate a bisection. Similarly, reasons are hyperintensional, and norms give rise to reasons. It is one thing to have a reason to bisect and another to have a reason to bisect-or-trisect. Finally, consider that if God exists, then, necessarily, p is true if and only if God knows p, since God is a necessary being that is essentially omniscient. But we are prohibited from bringing it

about that a murder is committed, while there is no prohibition on bringing it about that God knows a murder is committed.

The above examples all involve moral normativity. But plausibly the same is true of other kinds of normativity. The function of the $\times$ key on a calculator is to multiply quantities, not to calculate the exponential of the sum of their logarithms. It is the proper function of a duck embryo to develop two feet, but it is not the proper function of a duck embryo to grow a number of legs that God would believe to be the smallest prime number.

Note that on this normative account we don't need any causal connection to the objects we speak about. Mathematical objects are no more problematic than physical objects. Even if an infinite number of sets of mathematical objects satisfy the Peano axioms, it is open to the normative semanticist to say that there is a pair (N, s) that make it be the case that according to the norms of our nature saying "There are infinitely many primes" is appropriate just in case there are infinitely many members of N that are prime with respect to the successor function s, so that the members of N are *the* natural numbers. At the same time, the normative semantics could allow that there is no privileged system (N, s) but instead all mathematical statements are conditional. Settling the question of which of these is true is difficult to task for the philosopher of mathematics, but neither presents a special semantic difficulty.

2.1. Illocutionary force. Typically, one asks someone for something that one wants. But asking is not the same as communicating one's desire. First, sometimes one asks for something one doesn't want. For instance, a security specialist could conduct a phishing call where they ask a fellow employee for their password, hoping that few if any will give it. Or a middle manager might be tasked by upper management with requesting something from staff that the middle manager thinks is actually bad for the company, and hence hope that no one will agree to the request. Conversely, one may want something but not ask for it for moral reasons. To adapt a situation that occurs twice in P. G. Wodehouse stories??ref, one may own an ugly heirloom that one cannot give away because of one's relationship with the person from whom one received it, but one would be glad if it were

taken away. One could imagine a frank conversation where one happens to slip that one wouldn't mind the heirloom taken away. In the Wodehouse cases, the communication *is* a surreptitious request—and that, of course, is illegitimate, much as a king's exclamation "Would that someone rid me of this troublesome priest" is an invitation to murder. But one could also imagine a case where the slip is not a request, but simply a frank statement to a friend, followed by sincere emphasis that one isn't requesting removal. In that case, removal of the heirloom would be theft, even if it were desired by the owner. Or, for a different case, one might have a moral objection to a particular life-saving medical procedure, and hence one's conscience would forbid one from requesting it, but nonetheless wish that the procedure were done to one against one's will, say by a medical mistake, and one could in a frank conversation communicate that wish *without* that constituting an underhanded request.

In making a request, one creates a reason for the other party to provide one with someone. And not just any reason, but a special kind of reason in light of one's own request.???

But now consider the first time anybody ever requested anything. In requesting, they created a moral reason for their interlocutor. This was a power they already had, and the meaningfulness of the communicative act of requesting must have already been in place. How? How could that communicative act not only had its illocutionary force but been *understood* to have that illocutionary force given that no one had ever requested anything? The meaning of a request is largely defined by the kind of reasons it gives rise to. But how can one grasp these reasons if one has never encountered them before?

3. A sharp world

??backref to indeterminacy

CHAPTER VIII

Metaphysics

1. Composition

2. Ill-matched matter, rearrangement, the power to continue existing and immortality

3. Diachronic identity

CHAPTER IX

Laws of nature and causal powers

CHAPTER X

Evolution, Harmony and God

1. The origin of the forms

1.1. Evolution and forms. We have good empirical reasons to think that the variety of biological structures that fills our planet is largely or completely the product of unguided variation together with natural selection. However, as I have argued, there are good philosophical reasons to think that the organisms with these structures have normatively laden forms which specify how the organisms should behave, endow them with the causal powers that make that behavior possible, and impel them towards that behavior.

It is implausible to think that the forms supervene on the biological structures. For instance, one theory of the evolution of wings for gliding is that small wings are useful for heat dissipation. Larger wings allow for more dissipation of heat, but are also more expensive for the organism to maintain. However, at around size at which the heat-dissipation benefits are outweighed by the maintenance costs, the wings also become useful for gliding. It is plausible that a species A that has the smaller wings has them with the telos of heat dissipation. But a species B that has evolved the larger wings has them with the telos of gliding, either instead of or in addition to heat dissipation.??ref,check But we can now suppose a member of B whose wings are defective and only good for heat dissipation. Such a member's biological structure might be largely indistinguishable from that of a normal member of A, and yet it is normatively different: such wings are defective in B but entirely appropriate in A. If these norms are grounded in forms, it seems there is a different form in members of B than of A.

In general, in the evolutionary process, we expect small transitions in genetically-based biological structure between parents and children, with no change between the parent's

form and the child's form. For if we had constant change between the parent's form and the child's form, our best account would be that the form simply matches the biological structure, which would not allow for genetic defects, and yet genetic defects—deviations of genetically-based biological structure from the kind norms—are clearly possible. Moreover, it is important to our ethics that all human beings, despite a wide variety in physical and mental endowments—including the striking biological difference between male and female—are beings of the same kind.

We thus need an explanation of why it is that at certain apparently relatively rare and discrete points in the evolutionary sequence we have a new form on the scene. This itself yields Mersenne questions: while some transitions of form might happen to coincide with a particularly striking genetic transition, we expect a number of them to come along with only minor genetic transitions, seemingly at arbitrary positions. What explains these transition points?

Hitherto in this book, such questions were answered by invoking the forms themselves. And this can be done in this case as well. We might suppose that the form of species A endows the members of A with a causal power to generate new members of A in some circumstances, together with new instances of the form of A, but also a causal power to generate new members of B in other circumstances, along with new instances of the B form. The difference in circumstances could be determined by the DNA content in the gametes joining together, so that when a descendant is going to have such-and-such DNA contents, the descendant gets the form of A, but with other DNA contents, the descendant gets the form of B.

This story requires the forms to contain intricate specifications of which form is generated when. Granted, the slew of Mersenne questions we have already raised should make us circumspect about balk at mere complexity of form. But now observe that the story as given above requires that the first biological organism on earth—presumably some simple unicellular or maybe even proto-cellular?? organism—contain within it a form that codes for the causal power to produce forms of all possible immediate descendants of it. These

immediate descendant forms then would have to code for the causal power to produce all their possible immediate descendants, and so on. Thus, the form of the first and simplest organism would implicitly code for all the forms of life that would ever actually be found on earth, and indeed all the forms of life that *could* ever descend from it.[1] We thus have here a dizzying complexity.

???few species story!??? no help, still have complexity

But the problem does not stop here. For we can now ask where that immensely sophisticated form of the first organism comes from? If we say that it comes from the causal powers of non-living substances, such as fields or fundamental particles, then we have to posit an even greater complexity in the forms of these non-living substances. The result would be highly counterintuitive, by supposing non-living things to have immense sophistication of form. Further, however, we would need a story of where the first forms arose from. If we take the above account to its logical conclusion, then at the Big Bang we would already have particles or fields whose forms implicitly included the vast formal complexity of all physically possible living organisms. And this in turn yields a powerful design argument. For the idea that such complexity would simply come about for no reason at all is utterly implausible.

Thus, the story that forms contain the rules for the generation of future forms points towards a being whose own power is sufficiently great to generate such forms. And to avoid a vicious regress, such a being would need to be a necessary one.

But note that once we have accepted the existence of a necessary being that is the ultimate source of the varied forms in our world, we can now tweak the story to avoid the implausible idea that unicellular organisms implicitly code for the forms of elephants and unicorns. Instead of supposing that the transitions between forms corresponding to certain selected changes in genetic structure are caused by the parent forms, we can suppose that

[1] It is tempting to say that the number of possible descendant forms is infinite, but that is not clear. After all, there could be some physical limit to the size of the genetic code fo a biological organism given our laws of nature. But in any case, finite or not, the number of possible descendant forms is incredibly large.

the necessary being is directly responsible for the transitions of forms. On such a view, the form of a unicellular organism might only endow its possessor with the ability to generate a descendant of the same kind, and the necessary being would directly produce any new forms when it is appropriate to do so.

1.2. Reasons for creating forms. Of course, this would lead to the question of *why* the necessary being produces new forms when it does so. Here, taking the necessary being to be rational can help. For there can be good value-based reasons for the transitions to fall in some places rather than others.

Consider, first, an odd thought experiment. A horse-like animal comes into existence with an maximally flexible form such that whatever the animal does fulfills the norms in the form. To eat and grow is one proper function, and to starve and produce a corpse is just as proper a function. Whatever our "flexihorse" does or undergoes is equally good for it. But there is something unsatisfactory about the flexihorse as a creature. If whatever the flexihorse does is equally good for it, then the fact that the flexihorse flourishes is just a direct and trivial consequence of its externally imposed form rather than the individual's *accomplishment*.

Reflection on this suggests there is a value in creating organisms that can fail to fulfill their norms. This value might be grounded in the forms themselves: it might be that real horses, unlike flexihorses, have self-achievement of flourishing among the proper functions in their form. And there is a value in creating organisms that have additional types of good written into their form, including such self-achievement. Alternately, one might hold that in addition to kind-relative goods, there may be kind-independent goods—perhaps grounded in imitation of the creator??forwardref?—and self-achievement of flourishing could be one of these.

Either way, a rational being creating organisms has reason to create organisms that can fail to achieve their form, and hence has reason to create beings with less flexible norms. Moreover, there appears to be a comprehensible value—again, either kind-relative or kind-independent—in production of beings of the same kind. As an intuition pump here, think

of the *Symposium*'s idea that the yearning for eternity is exemplified in animal reproduction. Thus, we can give a value-based explanation for why a necessary being would create beings in discrete kinds, with norms that the beings need not live up to.

2. Explaining harmony by natures and evolution

[2]

2.1. Number of natures.

2.2. Nomic coordination.

2.3. Aristotelian optimism revisited.

2.4. Fit to DNA.

2.5. Fit to niche.

2.6. Nature zombies.

2.7. Exoethics.

2.8. Aquinas' Fourth Way and the good. Aquinas' Fourth Way??ref puzzles the modern reader. It begins with a principle that comparisons between degreed properties are grounded in a comparison to a maximal case: one is more F when one is more like the item that is maximally F. Aquinas then illustrates the principle with the case of heat and fire: an object is hotter provided that it is more akin to the hottest thing, namely fire. He then applies the principle to goodness, and concludes that there is a best thing, and this is God.

The fire illustration is not just unhelpful to us, since we know that fire is not the hottest thing (the sun is almost twice as hot as the hottest flame), but it is actually a conclusive

[2]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.

counterexample to the degree property principle, since we can easily compare temperatures without reference to an alleged hottest object.[3]

So Aquinas' comparison principle is false. But I contend that there is still something to his argument when applied to the good.

Now, a form-based metaphysics gives a powerful account of the good for a being of a particular kind—an oak, a sheep or a human, say—in terms of its match to the specifications of the form. It also gives a ground to comparisons between the good of different instances of the same kind: a four-legged sheep is, other things equal, better at sheepness than a three-legged sheep, because it more completely fulfills the specification in their ovine nature. In fact, this is itself a counterexample to Aquinas' comparison principle, in that we can compare degrees of success at sheepness without supposing any individual sheep to be perfect.

However, in addition to value comparisons within a kind, there are ones between kinds. When Jesus says that we are "worth more than many sparrows" (Mt. 10:31??ref), what he says is quite uncontroversial. Indeed, even a perfect sparrow seems to have less good than a typical human. While the nature of a sparrow will enable value comparisons between sparrows, and that of a human between humans, we still have the question of what grounds the value the difference between sparrows and humans. Some Aristotelians reject cross-kind value comparisons as nonsense.??refs But given the intuitive plausibility of many such comparisons, this rejection is a costly one.

Aquinas' Fourth Way is not infrequently seen as more Platonic than his other arguments for the existence of God, and Plato indeed had a solution to the problem of cross-kind comparisons, by talking of differing degrees of imitation of the Form of the Good, which itself is perectly good. Plato, on the other hand, lacked a satisfactory solution to

[3]In any finite universe, presumably there will be a hottest object. However, temperature comparison is not defined by that object, since even if Bob is in fact the hottest object, we would expect it to be physically possible to have a hotter object than Bob. But if degrees of heat were defined by closeness to Bob, it would not be possible to be hotter than Bob, since nothing can be closer to Bob than Bob.

the problem of intra-kind comparisons. He may well have thought that there was a Form of Humanity??refs, which exemplified humanity perfectly, so that similarity to the Form of Humanity would define how good one is at being human. However, we can see that this solution is clearly unsatisfactory. First, the Forms are immaterial, so the Form of Humanity is immaterial, and hence it lacks fingers. Thus, the fewer fingers a human has, the more they are like the Form of Humanity, and hence, absurdly, the more perfect they are. Second, if somehow the Form of Humanity ends up having body parts, then the Form of Humanity either has an even number of cells or an odd one. But clearly neither option is more perfect than the other.

Central to Plato's solution to cross-kind value comparisons is the self-exemplification of the Form of the Good: the Form of the Good is itself maximally good. But a similar self-exemplifying Form cannot be used to account for intra-kind comparisons. Aristotle, on the other hand, has the non-self-exemplifying forms immanent in things. The Aristotelian form of humanity specifies human perfection, but does not do so by exemplifying it. It has neither fingers nor cells, but it *specifies* that humans should have ten fingers while specifying an age-dependent normal range of cell numbers rather than a specific cell count.

Notwithstanding the general falsity of Aquinas' comparison principle for degreed properties, Aquinas provides us with a plausible extension of the Aristotelian system to allow for comparisons of degrees of good between objects of different kinds in terms of the similarity to or degree of participation in a maximally good being, a divine being that plays the role of a self-exemplifying Form of the Good. The human being participates in God in respect of abstract intellectual activity, Aquinas will contend, while sparrows do not, and in that important respect, at least, humans are more like God. On the other hand, the sparrow's movements approximate divine omnipresence better than the stillness of a mushroom does, and in that respect at least the sparrow is superior to the mushroom. We have, thus, a ground for something like a great chain of being.

There are still difficulties here. While the human is superior in intellectual activity, the sparrow moves around with greater three-dimensional freedom. How can we say that the

human is superior all things considered? Where we previously had a problem of cross-kind comparisons, we now have the problem of cross-attribute comparisons. Intuitively, the human's intellectual superiority to the sparrow trumps the sparrows motive superiority to the human, and enables us to say that the human is more perfect on the whole. This higher level question is difficult indeed.

But there is some hope in thinking that in attributing different divine attributes we sometimes express divinity to different degrees. It may be that there is no meaningful comparison between how well we express divinity by saying that God is all-knowing versus by saying that God is all-powerful, saying that God knows the multiplication table up to 10×10 expresses divinity less well than saying that God can create any possible physical reality. I suggested earlier that motion imitates divine omnipresence. Thus, the sparrow's ability to fly imitates God's presence throughout several kilometers surrounding the surface of the earth, while the human's more limited mobility imitates God's presence in a thin two meter shell of air surrounding that surface. But the degreed difference between the two divine attributes—each a limited special case of omnipresence—imitated here might well be trumped by the fact that the sparrow does not imitate God's abstract intellectual activity *at all* while the human does imitate that activity, and does so in respect of a very wide scope of things (the human can think abstractly about the whole universe, for instance).

In ??backref, we gave a non-theistic Aristotelian sketch of a three-step great chain of being. The account here has a hope of allowing one to fill in more intermediate links. Even if the details in the comparisons between different attributes or respects do not work out, we still have an advantage for the theistic Aristotelian in being able to make cross-kind comparisons under specific respects, like motility or intelligence.

??ref:Jeffrey/Ward

2.9. 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???]

2.10. Ethics and happiness.

2.11. Norms that fit with modern technology and any real but outlandish scenarios. In ??backref, I argued that an ethics based on human form can simply ignore outlandish scenarios that are far outside of our ecological niche, such as ones involving infinite numbers of beneficiaries. However, there is a danger in this line of reasoning. As Arthur C. Clarke famously said, "Any sufficiently advanced technology is indistinguishable from magic."??ref To human beings 50,000 years ago (or even just 500 years ago!) much of our technology would indeed be magical, and decisions that we routinely need to make, say in bioethics, would be predicated on outlandish assumptions.

We might thus expect an ethics and epistemology grounded in a form possessed by hunter-gatherer primates to be silent on dilemmas of a highly technological society, leaving us to do whatever we wish, or, even worse, to fail to harmonize with the shape of our lives, like that of a fish on land. Yet while there are, as there have always been, difficult and controversial moral and epistemological cases, we do not in fact find ourselves adrift without guidance in the modern world. Virtue continues to contribute to our flourishing, and ancient texts, whether religious or philosophical, continue to point to good ways of living.

This gives us reason to think that if our moral norms are grounded in human nature, human nature was somehow picked out with foresight for what kinds of challenges humans would face in the distant future. Our ethics does may not work in outlandish situations, such as those involving infinities as noted in ??backref, but it works in a broader range of moral environments much broader than that found in early homo sapiens society. Thus, the theistic version of our natural law theory both accounts for the apparent unsatisfactoriness of our ethics in situations that humans apparently never find themselves with and the applicability in situations across a very wide range of situations, wider and more technologically varied than the natural environment of other animals. This kind of foresight points to a foreseer, indeed a designer, and hence towards a theistic version.

The move I suggested in ??backref for outlandish scenarios, namely that our ethics and epistemology simply does not apply to them, may be problematic, however. For *some* seemingly outlandish scenarios could turn out to be real. Many religious people think that some or all individual humans will live forever. And many people, religious or not, think there is a serious possibility that human beings will spread through the galaxy, affecting vast numbers of lives, which may make actual seemingly outlandish questions about where our actions have very slight probabilistic effects on vast outcomes.??ref:Fanaticism Here, a theistic story might also be attractive. God can know what kinds of seemingly outlandish scenarios are actually relevant to the lives of his creatures, and can wisely choose the forms whose norms that fit with these. The resulting norms may seem *ad hoc* at times: they won't be the elegant principles of classic utilitarianism, for instance, but those kinds of principles face great difficulties in outlandish scenarios. But a wise rule does make judgments that can be *ad hoc*.

2.12. Avoiding radical scepticism.

3. Explaining harmony theistically

4. Explanations of moral norms

4.1. A pattern of ethical explanation. Here is a familiar pattern. We have a deeply-seated moral intuition about the general prohibition, call it g, of some action, such as incest. It is not clear how to derive the prohibition from intuitively more basic principles, such as one of the categorical imperatives. Easy considerations, which I will call the cs, show that in *typical* cases the action is wrong, but our moral intuition goes beyond these typical cases. Thus, considerations of the abuse of power, distortion of familial dynamics, and genetic harms show that most cases of incest are wrong, but it is easy to imagine cases of incest to which these considerations do not apply—say, elderly siblings who were raised apart—and yet moral intuition forbids incest in those cases as well.

We can now save the moral intuition by saying that the more general prohibition g is simply a fundamental moral rule, not reducible to the cs that explain why the action is

wrong in typical cases. But if we stop at this, the connection between g and the cs mere happenstance, and that seems intuitively wrong. The abuse of power, distortion of family dynamics, and genetic harms should be relevant to why incest is wrong.

At this point, often we are in a position to see another fact: it is quite beneficial to have a general moral prohibition beyond the prohibitions arising from the cs.

One reason for such a benefit from a general prohition could be that our judgment as to whether the cs apply to a given case is fallible, especially given our capacities for self-deceit, and the costs of violating the cs are so high that it would be better for us to have a general prohibition than to try to judge things on a case-by-case basis.

Second, in some examples of the pattern, serious deliberation about the forbidden action can itself harm one or more of the goods involved in the cs: thus, having to weigh whether the distortion-of-family-dynamics consideration applies against a particular instance of incest can itself distort the agent's participation in family dynamics.

Third, we could have a tragedy of the commons situation. It could be that the cs are actually insufficient to render an instance of the action wrong, but we would be better off as a society if we had general abstention from the action. Thus, perhaps, the genetic harm coming from one more couple's engagement in incest would be insufficiently significant to render the incest wrong, but without a general prohibition, incest would be sufficiently widespread as to cause serious social problems. A general prohibition that is not logically dependent on the cs would help avert such social harms.

These considerations are very familiar to us in the case of positive law. Jaywalking involves harms such as disruption of traffic flow and the danger of death of the pedestrian and of trauma to the driver, and the considerations of these make jaywalking wrong in typical cases. There are obvious instances, however, where these considerations do not apply: say, crossing a road where the pedestrian can clearly see that there are no intersections or cars on the road for a significant distance in either direction. However, it may be better for people simply to abstain from jaywalking than judging whether the disruption and safety considerations apply on a case-by-case basis, because there could be so much harm if the

judgment were to go wrong. As a result, it is can be reasonable for a state simply to ban jaywalking altogether (or to ban it with some clear and easily adjudicated exceptions). We similarly resolve cases of tragedy of the commons with positive law: think, for instance, about laws against littering.

In the case of positive law we have two different explanations. First, there is an explanation of why the forbidden action is wrong in general: this is because it has been competently forbidden by legitimate authority. This explanation need not make reference to considerations such as disruption of traffic flow or danger of death.[4] Second, there is an explanation as to why the action has been forbidden by the authority—and here all the rich considerations are relevant.

A theistic version of natural law can have precisely this pattern. An action is morally forbidden because our nature is opposed to it. This explanatory fact does not make reference to the *c*s. But we still have a further question to ask. The most obvious way to ask the question is to query why our nature includes this prohibition. But since our nature is essential to us, the answer to that question could simply be the necessary truth that we couldn't exist without this nature. However, we can put the question in a slightly different way: Why are there intelligent primates on earth with a nature that includes this prohibition rather than some other kind of intelligent primates with a nature that does not include this prohibition? And here the theist can answer: Because it would be good, in light of the *c*s and the further considerations in favor of generalizing the prohibition beyond the cases where the *c*s specifically apply, to have intelligent primates with a nature that includes this prohibition, and God acted in light of this good.[5]

[4]Though in some cases *some* such reference may be needed in order to establish that the matter falls within the competence of the authority in question. Thus, a government agency may be permitted to make rules on matters where traffic flow disruption is concerned.

[5]A divine command theorist can make the same move, but divine command theory has some liabilities which were discussed in ??backref.

4.2. Global aesthetic-like features. [6]

4.3. Family.

4.4. Retributive justice.

4.5. Divine authority.

5. Kind-independent goods

Diversity, flourishing, self-achievement, etc....??? Imitation?

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

CHAPTER XI

Eternal Life and Fulfillment

??interact with Oderberg on suffering and pain

CHAPTER XII

Aristotelian Metaphysical Details

1. Introduction

2. Teleological reductions

2.1. A multiplicity of concepts. The applications above, and the Aristotelian tradition, make use of various normative concepts that are said to be grounded in forms, such as proper function, teleology, and flourishing. It would be good to investigate if these can be further unified, under either one of these concepts, or some further unifying concept.

First, observe that we have both binary and comparative normative concepts, often in the same context. Suppose a stranger is about to be hit by a train, and the only four options are:

(1) give them a kick to ensure that they have no chance of survival

(2) make fun of them

(3) stand by idly

(4) jump in and push them out of the way likely at the cost of one's own life.

The binary distinction is that the first two options are impermissible, and the other two are permissible. But there is a comparative distinction as well: (1) is worse than (2), and (4) is better than (3). The binary distinction does not reduce to degree of comparison: while (2) is much worse than (3), (3) is much worse than (4) (though it is more natural to say that (4) is much better than (3)). And the comparisons do not reduce to the binary characterizations.

Proper function and teleology appear to be primarily binary concepts: a thing functions properly or improperly, and a thing either does or does not achieve its end. Flourishing, on the other hand, seems to comprise both the binary and the comparative. The mildly

vicious person unjustly suffering horrendous pain appears not to be flourishing *simpliciter*, but if the pain were increased, they would languish more. Flourishing thus appears the best candidate for a foundational concept for our norms.

But because the notion of ends and teleology has been so important in the Aristotelian tradition, both in the case of voluntary action and involuntary activity, it is worth thinking some more about ends.

2.2. Ends. Many activities seem to occur for an end. The activity then counts as successful provided that the end occurs and occurs as a fulfillment of the activity. An organism produces gametes in order to reproduce. A cat chases birds in order to catch them, and eats them for nourishment. And I put on shoes to keep my feet comfortable when I walk. The end-directedness of much voluntary activity is obvious, but whether there really is teleology in the involuntary cases is more controversial, though highly intuitive.

The Aristotelian tradition tends to analyze voluntary action as always end-directed, but also tends to see involuntary activity as often, if not always, directed at an end. I will argue, starting with the case of voluntary action, that many interesting phenomena would be misclassified as end-directed. The actual structure can be more complex, and while it has a directional structure, it is misleading to think of that structure as teleological in the sense of possessing a *telos*, an end that fulfills it.

Consider a sprinter who is running a hundred meters all out against a clock, rather than against other opponents. The runner has an end, namely to sprint 100 meters. But sprinting 100 meters does not explain the intense effort the runner puts in. Less than half of the effort could have been put in, and 100 meters would still have been sprinted. The bulk of the runner's effort is explained by being directed not at completing the sprint but at completing it in minimum time.

But what state of affairs does the runner's speed-directed effort have as its end? A runner might have a particular target time in mind. However, we are imagining a runner who runs all out. A runner who is just aiming at a particular time could slow down if it

became obvious that a slower run would still achieve the target time, but not so our all-out runner. Our sprinter may have some specific time in mind to motivate themselves, but interpreting their action as merely aiming at that time does not capture all of the directional structure of the performance. Any shortening of the time of the run is welcome given the sprinter's aims.

We would normally describe the runner as trying to run 100 meters "as fast as possible", and that seems to be a coherent description of an end. However, the language of "as fast possible" should not be taken literally. First, we have the question of what the relevant comparison class is. Is the runner trying to run as fast as any human being can on any track? As fast as they themselves can run on this track on this occasion? Or something in between? ??? unlikely success!

Second, suppose we fix a particular sense of "as fast as possible", and then after ten meters the runner realizes that they have been slightly slower than is possible. At this point, it is no longer possible for the runner to achieve the goal of literally running the run as fast as possible. But we do not expect the runner to stop. We expect the runner to resume running all-out, as part of the same directed activity.

3. 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.

3.1. 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.

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