5

Humanity

In Kantian ethics, the fundamental value is humanity or rational nature as an end in itself. This value grounds the supreme principle of morality from the side of its matter and results in the moral law's second formulation (FH). In the most complete or universal formula of the moral law, FA, this value is developed into the ground of moral legislation itself, in the form of the dignity of rational nature as universally legislative. This chapter tries to explain the nature of this fundamental value and look at Kant's defense of the claim that rational nature has such a value. It concludes with a discussion of the possibly controversial moral status of some human beings or forms of human life, and of nonhuman animals.

1. What Is an End in Itself?

An existent or "self-standing" end. Rational nature is described as an end – an end in itself. Kant calls it a "self-sufficient," "independent," or "self-standing" (*selbständig*) end, in contrast to an "end to be produced" (G 4:437). It is an end in the sense of something *for the sake of which* we act. This is not a technical or "funny" sense of 'end'. It is simply the most basic and encompassing sense of the word. Rational nature is *not* an end, however, in another (more derivative) sense in which Kant also thinks that every action must have an end. In the claim that rational nature is an end in itself, rational nature is not being thought of a state of affairs to be produced by action. Instead, an "end in itself" is something already existing whose value grounds even our pursuit of the ends produced by our actions. The notion that the word "end" may refer only to such a producible state of affairs is simply a philosophical error about the concept "end."

Every moral action must have an end to be produced, but such actions must be grounded on a "self-standing" end. This is a direct consequence of the fact that this value is to motivate obedience to a categorical imperative – a principle that rationally constrains us without presupposing any end to

be produced. The value for whose sake we follow a categorical imperative cannot be the value of any end to be produced. From this Kant infers that it must be the value of something already existing whose value is fundamental and unconditional.

As was mentioned in Chapter 4, critics of Kantian ethics sometimes complain that the concept of a categorical imperative makes no sense because there could be no reason for obeying such an imperative. This is usually because they think that the only reason for obeying an imperative must be an end in the sense of an end to be produced. They do not notice that Kant's concept of an objective end in itself is precisely his answer to their question. The conceptual features of an end in itself that we have noted follow simply from the fact that it must provide the rational motive for obedience to a categorical imperative. If there are categorical imperatives, Kant reasons, then there must be a reason for obeying them, and such a reason can consist only in something that is an end in itself.

An objective end. In order to ground a categorical imperative, the end in itself must have another distinctive property: It must be *objective* – that is, valid for all rational beings irrespective of their inclinations. Acting for its sake must constrain rational volition without presupposing or depending on any contingent empirical desire of the willing being. This does not mean that action on it must be action from which desire is absent. Rather, in accordance with the end in itself, pure reason of itself produces desires – desires for the ends to be produced that are set in accordance with the objective value of the end in itself.

'Respect' is the name for the proper attitude toward any objective value.[1] Depending on the nature of such a value, it may call for widely varied kinds of conduct. Some objective values are to be promoted, while others are to be exemplified, appreciated, or honored, or simply not violated. Sometimes the maximal promotion of a value involves its violation, so priorities among these different kinds of conduct sometimes matter. We saw in Chapter 3 that cases can arise where the promotion of a value might come into conflict with the exemplification of the same value (as when being tolerant might require us to permit someone to preach intolerance). An ethical theory that considers only the consequences of actions for the promotion of values will sometimes go far wrong if it countenances the violation of the very value promoted.

To say that we act for the sake of something already existing does not mean that we act for the sake of *bringing about* its existence or *preserving* its existence. For that existence is merely another *state of affairs*, another possible end to be produced (which would mean that the moral imperative is hypothetical after all). If it is normally a requirement of morality that we should seek to preserve rational beings in their existence, then this is a *consequence* of the fact that if an existent being has basic and unconditional value, then the state of affairs of its continued existence also has great value,

at least most of the time. But from the fact that humanity or rational nature has dignity, or fundamental and unconditional value, it by no means follows that the value of human *life* is basic or unconditional. At times people are in terrible situations where living up to the dignity of their rational nature even requires them to sacrifice their continued existence. There may also be situations in which moral rules grounded on the worth of rational nature as end in itself require that human beings be killed, or even entail that the continuation of a human life should no longer be set as an end at all. FH, as a formula of the supreme principle of morality, is consistent with all these possibilities and cannot all by itself determine how often or how seldom they will occur. For better or worse, Kantian principles (rightly understood) justify attaching great importance to preserving human life, at least most of the time, but they provide no support for the idea that, as some people like to put it, "all human life is sacred."[2]

Treating a being as an end in itself means respecting the value of what makes it such an end. After we see that this value resides in rational nature, we see it implies that, at least in general, rational beings should not be subjected to deception or coercion. Instead, we should seek to harmonize our strivings with those of other rational beings toward their ends. FH thus naturally leads toward the ideal of FRE, in which the ends of all rational beings would ideally constitute a systematic combination or "realm."

"Not merely as a means." Much is sometimes made of Kant's claim that we must treat humanity as an end, *never merely as a means. Far too much*, in fact. One fallacious pattern of reasoning begins with the proposition that a person is being treated as a means and concludes merely from this that they are not being treated as an end in itself. But it is possible to treat persons as ends in themselves and also as means, as long as you respect their rights and dignity. This is not only possible, but Kantian ethics positively enjoins it. FRE tells us to obey the laws of a realm of ends. A *realm* is a combination of rational beings whose ends harmonize and all of whose actions serve as means to a systematic combination of ends. In a realm of ends, every rational being would therefore be treated as an end in itself and at the same time as a means to this system of shared ends.

It is also fallacious to infer solely from the fact that someone is not being treated merely as a means to the conclusion that they are being treated as they ought to be under FH. In sympathetic depictions of the abominable American institution of slavery, some white masters (such as the genteel Ashley Wilkes in *Gone with the Wind*) are shown caring about the welfare of their black slaves, which shows that they did not treat them *merely* as means. These supposedly exemplary masters, for all their evangelical Christian benevolence, nevertheless fail monstrously to treat their slaves as ends in themselves. Kant himself makes a similar observation about the lame excuses offered by a feudal landlord who treats his serfs with paternalistic benignity (MS 6:454).

2. Humanity Is an End in Itself

If there is to be a categorical imperative, then something must be an existent, objective end in itself. The question is: What? Kant's claim is that the sole end in itself is "humanity" or rational nature in persons. What is meant here by 'humanity'?

Humanity as a predisposition. According to Kant, our nature has three fundamental "predispositions" (*Anlagen*): animality, humanity, and personality (R 6:26). Animality contains our instinctual capacities for the survival of the individual and the species: "mechanical" (prerational) self-love (self-preservation), sexuality (preservation of the species), and the social drive – our instinctual need to be in community with other human beings. Humanity contains our rational capacity to set ends and devise means to them, and our rational self-love, giving us grounds for forming a conception of our happiness and pursuing it. Personality is our rational capacity to legislate for ourselves the moral law and obey it.

Sometimes Kant distinguishes within "humanity" between two different predispositions: the *technical* predisposition to devise means to arbitrarily selected ends, and the *pragmatic* predisposition to rational self-love, which specifically involves our sociability as rational beings (VA 7:322–4). It includes the ability to use *other human beings* as means to our ends (placing this means–ends relationship in a different category from the technical one involving the use of things) and also our capacity for culture or self-perfection, the development of new ways of thinking and modes of life, which again Kant treats as different from a merely technical relationship to skills or instrumental mechanisms). The technical and pragmatic predispositions, of course, correspond to technical (instrumental) and pragmatic (prudential) rationality, and personality corresponds to moral rationality, as we distinguished the three norms of reason in Chapter 1, §4.

It is noteworthy that what Kant claims to be an end in itself, possessing the absolute objective worth that grounds our obedience to moral laws, is *humanity* – especially in this last (pragmatic) sense. It is not animality or even the technical (instrumental) rationality that has this value. Nor (perhaps more surprisingly) is it our *moral* predisposition (though Kant holds that it is that predisposition which gives us *dignity*). The absolute worth that grounds morality is the predisposition toward prudence (rational self-love and the end of our own happiness), rational social interaction, and the cultivation of ourselves and all our faculties through society in the course of human history.

Arguments for an ultimate value. Kant's arguments for the bold thesis that humanity is an end in itself are terse and obscure. The claim that rational nature in persons has this status, however, exercises a powerful influence on modern moral thinking not only in philosophy but even in ordinary life and in moral common sense. We may have to face the fact that the mere

claim that human beings have absolute worth as ends in themselves may in the end be more compelling all by itself than any argument that Kant or anyone else could ever offer for it. But for a proponent of the kind of ethical theory sought by Kant or Mill, this fact can never be satisfactory all by itself to justify a basic value.

As I mentioned in Chapter 3, Kant's task at this point is similar to (though obviously not the same as) the task J. S. Mill sets himself in *Utilitarianism*, where he proposes to show that happiness is desirable and even the only thing desirable for its own sake. I have always thought that Mill's argumentative strategy, and at least crucial parts of his argument, is defensible against the common criticisms. Kant's conclusion goes deeper than Mill's, enabling us to explain, in a way that Mill cannot, *why* happiness is rationally desirable. But I think both philosophers do about as well as it is possible to do in arguing for an ultimate value.

According to Mill, the only argument to which claims about ultimate value are susceptible is one that shows this value to be one we already acknowledge both in theory and in practice (Mill p. 35). The strategy of such an argument is to cite what we do, and what we must represent ourselves as thinking and doing, when we form preferences, set ends, and make decisions, and then to argue that these actions, thoughts, and representations are best understood as recognizing something as an ultimate value. The phrase "best understood" should not be taken as claiming there are no logically possible alternative understandings – to claim that would be to set an impossibly high standard for this kind of argument – but only as the claim that ascribing to ourselves this judgment of ultimate value is the most reasonable way of understanding what we are doing and thinking.

Perhaps, therefore, no argument about ultimate value can be expected to convince everyone – there are simply too many possible views about what is ultimately valuable, and too many clever philosophers too firmly attached to their own peculiar notions to expect that even a clearly more reasonable interpretation of their conduct will be able to convince them. (It should also be admitted, however, for pretty much the same reason, that no philosophical argument about *anything* can be expected to convince *everyone*.) But it helps when that claim of ultimate value is one that many people, perhaps even most people, are prepared to accept even without argument. If, in addition to this, the argument shows that this claim about ultimate value is a reasonable interpretation of what we are committed to in our thinking and doing, then it has done everything we should ever expect of it.

Kant has begun his argument already in the same spirit, by presenting moral obligation as grounded on categorical imperatives, and then showing that we can be motivated to obey a categorical imperative only if there is something that is an objective end in itself. He proceeds by eliminating some candidates for what the end in itself might be. He first argues

that the objects of our inclinations cannot be objective ends in themselves because their value is subjective and conditional on our having these inclinations.

About our inclinations ourselves Kant makes a claim that shocks many people: He says that inclinations are so far from being ends in themselves that we rightly regard them as a burden, and it would be rational to wish ourselves entirely free of them (G 4:428). This is a position he often associates with the ancient Cynic school in ethics (VE 27:248, 29:604), which he regards as the least plausible of the ancient schools. We capture Kant's real views here if we expand what he says a little: "Our inclinations themselves are so little of absolute worth, to be wished for in themselves, that there was even an influential school of ancient ethics that taught that the best means to happiness was to be entirely rid of them." This does not mean that Kant actually agrees with the Cynic view: In later writings he says it would be not only irrational but even immoral to wish to be rid of our natural inclinations (R 6:57–8). What Kant needs to show here does not require his coming anywhere near agreement with the Cynics. He needs only to claim that we do not regard any of our inclinations as objects of respect or as objectively and unconditionally valuable independently of the possible rational value of their objects. And this much seems obvious. Finally, Kant invokes the distinction between things and persons, claiming that only the latter, not the former, are ends in themselves (G 4:428). This supports Kant's claim only by suggesting – quite reasonably, I think – that it might be the best explanation for this distinction and the practical use we make of it.

The worth of humanity as a necessary presupposition of rational volition. Kant then presents his principal argument that rational nature is the end in itself:

> The human being necessarily represents his own existence [as an end in itself]; thus to that extent [FH] is a *subjective* principle of human actions. But every other rational being also represents his existence in this way consequent on just the same rational ground as is valid also for me; thus it is at the same time an objective principle from which, as supreme practical ground, all laws of the will must be able to be derived. (G 4:429)

What does Kant mean in claiming that *every* human being *necessarily* represents his own existence as an end in itself? He cannot possibly mean that, as a matter of contingent empirical fact, all people actually assent to the proposition that their own existence is an end in itself, in the abstract and somewhat technical sense in which Kant has just introduced the notion of an "end in itself." Most people probably never even entertain that proposition. Nor could that contingent empirical claim possibly be one he is making, because his assertion is that human beings *necessarily* represent their existence in this way. When he says that FH is to that extent "a subjective principle

of human *actions*," it is more reasonable to interpret him as meaning that there is something in the way that people act, and think about their action, which *necessitates* (or commits them to) representing their own existence as an end in itself. The question is: What could that be?

Kant holds that the most basic act through which people exercise their practical rationality is that of setting an end (G 4:437). To set an end is, analytically, to subject yourself to the hypothetical imperative that you should take the necessary means to the end you have set (G 4:417). This is the claim that you rationally ought to do something whether or not you are at the moment inclined to do it. It represents the action of applying that means as *good* (G 4:414) – in the sense of "good" that Kant explicates as: what is required by reason independently of inclination (G 4:413). Kant correctly infers that any being which sets itself ends is committed to regarding its end as good in this sense, and also to regarding the goodness of its end as what also makes application of the means good – that is, rationally required independently of any inclination to apply it. The act of setting an end, therefore, must be taken as committing you to represent some other act (the act of applying the means) as good.

In doing all this, however, the rational being must also necessarily regard its own rational capacities as authoritative for what is good in general. For it treats these capacities as capable of determining which ends are good, and at the same time as grounding the goodness of the means taken toward those good ends. But to regard one's capacities in this way is also to take a certain attitude toward *oneself* as the being that has and exercises those capacities. It is to *esteem* oneself – and also to esteem the correct exercise of one's rational capacities in determining what is good both as an end and as a means to it. One's other capacities, such as those needed to perform the action that is good as a means, are also regarded as good as means. But that capacity through which we can represent the very idea of something as good both as end and as means is not represented merely as the object of a contingent inclination, nor is it represented as good only as a means. It must be esteemed as unconditionally good, as an end in itself.

To find this value in oneself is not at all the same as thinking of oneself as *a good person*. Even those who misuse their rational capacities are committed to esteeming themselves as possessing rational nature. It also does not imply that a more intelligent person (in that sense, more "rational") is "better" than a less intelligent one. The self-esteem involved in setting an end applies to any being capable of setting an end at all, irrespective of the cleverness or even the morality of the end setting. Kant's argument supports the conclusion, to which he adheres with admirable consistency throughout his writings, that all rational beings, clever or stupid, even good or evil, have equal (absolute) worth as ends in themselves. For Kantian ethics the rational nature in every person is an end in itself whether the person is morally good or bad.

Kant's argument also does not involve saying (as Korsgaard wants to) that setting an end *confers* value on the end.[3] On the contrary, setting an end is an exercise of *practical reason* only to the extent that we think there is already *some good reason* for us to set that end. The value of the end is to be located in that reason, which must have existed already prior to our rational choice. Of course, if it is true that the sole fundamental and unconditional value is the value of rational nature as an end in itself, then the goodness of any other end must somehow be grounded in this value. Ends to be produced will usually have value, for instance, because they fulfill the needs, or enrich the lives, or contribute to the flourishing and the happiness of rational beings, and so setting and achieving these ends shows respect and concern for the value of those rational beings. But to hold that the worth of other goods is derivative from or dependent on the worth of rational nature in this way is not at all the same as saying that they have their goodness *conferred* on them by the choices of rational beings. On the contrary, we choose these other goods because they fulfill our needs or contribute to our happiness. It is not the case that our choosing them *brings it about* that they fulfill our needs or make us happy. Still less should we say, as Korsgaard also has, that rational beings *confer on themselves* the value of being ends in themselves.[4] Kant's claim, as I understand it, is that we necessarily regard rational nature as an end in itself *objectively* and *unconditionally*. Its being an end in itself could therefore not be contingent on any act of ours through which that value might be conferred. Rather, the argument is that it is our basic act as rational beings, the act of setting ends and regarding them as good, that necessitates our representing ourselves as *already* ends in themselves.

Generalizing to all rational beings. The next step in Kant's argument is to claim that every other rational being represents its existence as an end in itself through the same rational ground that is valid for me. Hence FH is not merely a subjective principle but also an objective principle. Not only my rational nature, but the rational nature in every person, is an end in itself. This inference seems correct if the rational ground for regarding myself as an end in itself is the capacity to set ends. For that capacity does belong to every rational being as such, to others as well, and, once again, to stupid and wicked people exactly as much as to clever and virtuous ones.

To this claim, however, Kant appends a curious footnote, saying that it is set forth here as a "postulate," the grounds for which will be given in the next section. It is not immediately clear, however, what grounds in the Third Section Kant has in mind, or even what claim Kant is saying is being presented for now as a postulate. My best guess is that Kant's provisional "postulate" is that my representation of myself as an end in itself is based on a *rational ground* – that is, that it is not merely a contingent fancy or cobweb of my brain. For in the Third Section grounds for this are presented, in the form of the freedom of the will, which (Kant argues) is presupposed

by practical reason and even by theoretical reason (G 4:447–8).[5] If that conjecture is right, then the argument here in the Second Section is that if there is such a ground (which turns out later to be freedom of the will and is for the moment merely postulated), then it must hold equally for all rational agents, who are therefore all equally ends in themselves.

It also makes sense to consider this claim a "postulate" in something like the Euclidean sense, where the Greek word for 'postulate' (*aitema*) means request. A postulate involves a request to perform an action (e.g., drawing a straight line between two points) and then also a request to grant some proposition on the basis of that action (e.g., that between any two points such a line can be drawn). At this point in the *Groundwork*, the requested action is that of setting an end, where this involves treating the end and the necessary means as good; and the proposition to be granted is that this action presupposes that there is a rational ground for regarding yourself, as a being having a capacity to do all this, as having unconditional and objective value as an end in itself.

Limits of the argument. Kant's argument does not work by showing that rational beings *are* ends in themselves but only by showing that in setting ends according to reason, *we must presuppose* that they are. But the argument also does not show that there is *no conceivable alternative* to representing oneself – and therefore every other rational being as well – as an end in itself. The setting of ends and the use of means to them might be understood, for example, not as a rational process but as a merely mechanical causal one, as Kant thinks it actually is in the instinctive teleology found in the behavior of nonrational animals. Or the representation of something as an end might be taken as a merely theoretical act of perceiving the goodness of an object, a passive state that would move us of itself, rather than an act of rational judgment carrying with it a practical authority for us that is worthy of esteem as an end in itself. No doubt there are still other alternative conceptions of our agency that philosophers might devise that do not support the commitment to represent one's rational nature as an end in itself.

Kant's argument, therefore, cannot and need not rest on the claim that all these alternatives to his interpretation of rational action can be conclusively refuted. It involves only the claim that his interpretation is more natural and reasonable than they are. I also think that so understood, Kant's argument does as much as can possibly be required of any argument purporting to establish a claim about what has ultimate value. In philosophy, as Aristotle wisely tells us, we must not apply the wrong standards to a subject matter (Aristotle 1094b25). This also means we must not expect more of a claim, or an argument for it, than is reasonable. When we ask the impossible, ignoring an argument's real but necessarily limited accomplishments, we will find the argument unsatisfactory, but that is our fault, and not a defect in the argument.

3. The Dignity of Humanity

In combining FUL with FH and advancing to FA, Kant makes a further claim about the moral status of rational nature in persons: He claims that it has "dignity" (*Würde*). The traditional meaning of this term involved identifying certain classes of people possessing a determinate social status that makes them superior to others. We have now perhaps become accustomed to Kant's extension of the term to all human beings, but we should not fail to hear in the phrase Kant's defiant and paradoxically egalitarian assertion that the highest possible worth any human being can have consists in a value that all human beings have equally – whether well born or ill born, rich or poor, intelligent or stupid, even good or evil. This radical egalitarianism, grounded in the conception of every human being as a rationally self-governing agent, is the most fundamental idea in Kantian ethics. The potential of this Kantian idea to transform our relations with one another is still pitifully far from being realized, or its implications even properly thought out consistently.

Dignity and price. That which has a "price" may be rationally sacrificed or traded away for something else whose price is equal or greater. That which has *dignity*, however, has a value that may not be rationally traded away or sacrificed, not even for something else that has dignity (G 4:434). In that sense, its value is *absolute*.[6] One conclusion that immediately follows from this is that respecting the dignity of one person cannot ultimately conflict with respecting the dignity of another. Thus the ends involved cannot ultimately conflict but must constitute a systematic combination or realm. This leads us from FA to its more intuitive variant FRE.

If being an *end in itself* constitutes the worth of *humanity* – in the technical Kantian sense, which is the capacity to set ends according to reason – then having *dignity* constitutes the worth of *personality* – which is the capacity to give oneself moral laws and obey them. Kant nevertheless frequently speaks of the "dignity of humanity" as well as the dignity of personality. Kant usually writes as if humanity and personality are necessarily coextensive.[7] I think they *are* necessarily coextensive. For setting ends according to reason is an act of freedom – involving at least freedom in the negative sense, because no impulse or inclination can necessitate my setting its object as an end (MS 6:381). But Kant holds that the concept of positive freedom, the capacity of giving oneself laws and having a reason that is of itself practical, flows from that of negative freedom, as constituting the essence of negative freedom (G 4:446, KpV 5:33); conversely, the capacity of positive freedom clearly entails the capacity to set ends according to reason.

Kantian ethics rests on a single fundamental value – the dignity or absolute worth of rational nature, as giving moral laws and as setting rational ends. The fundamentally valuable thing in the universe is a rational being, a person – or, more precisely, rational nature in a person. The demands made on us by this value depend on the kinds of conduct required to show

respect for this value. Other things having objective value have it, in one way or another, on the ground of this basic value. For example, it grounds the value of human happiness, and also of the perfection of talents people choose to develop. Some things that people rationally choose to make their ends acquire greater objective value because they rationally choose them. If you choose to develop one talent rather than another, then others have a reason to help you develop the chosen talent that would not exist but for your choice. Yet this example represents only a special case, and even here you could *rationally* choose to develop your talent only if you recognized it as already having some objective value (as the perfection of a rational being), which your choice could not possibly have been conferred on it. The idea that any objective value could be simply *conferred* by human choice is nonsense – it contradicts the very concept of objective value.

4. The Personhood of Human Beings

Who are persons? I have claimed that in holding rational nature to be an end in itself, and to have dignity, Kantian ethics articulates an idea that is widely appealing and fundamental to modern moral consciousness. But in the precise form I have just expressed it, this idea might also be seen as having certain consequences that are paradoxical, if not objectionable. The idea seems to grant fundamental moral status solely to *persons* – that is, to rational beings who are capable of instrumental, of prudential, and above all of moral reason, and who are morally responsible for what they do. (Let's call such beings persons *in the strict sense.*) It might be thought that other beings, such as children who are not yet persons in the strict sense, or even nonrational animals, also have moral status, a claim on moral concern, even certain rights. Don't children have the same rights to life and equal concern as adults? Don't we have moral reasons to concern ourselves with the welfare of nonrational beings, such as animals? Mustn't that status rest on some value independent of the rational nature in persons?

Kantian ethics must answer the last question in the negative, but it answers the other two in the affirmative. I think the right account of the moral status of nonrational living things and of human beings who lack personality in the strict sense can best be derived from Kantian principles, even though Kant himself did not worry about these questions as much as he should have, and some of the things he said about them do not seem to me entirely cogent, or to be the best account available to him.

Let us begin with the moral status of children or other human beings who at least temporarily lack the rational capacities constituting personality in the strict sense. In discussing family right, Kant declares that children are persons and treats them as having pretty much the same status as adults (though he would not grant them the right to direct their own understanding unaided by an adult guardian until they reach the stage of life at which

they are capable of this). Kant does not grant them the capacity to set ends according to reason or regard them as morally responsible for their actions, yet he never explains why for most purposes they should be treated as persons.

"Unity of the person." One approach to this topic relies on what its proponents like to call the "unity of the person."[8] They claim in effect that a human being is the same being at all stages of its existence, including those in which it is not yet (or no longer, or temporarily not) a person in the strict sense, and the Kantian view should be that such a being has the same moral status (the same rights, etc.) at all stages of its existence. The idea that every human being has the same moral status at all stages of its existence is an intuitively appealing one. The question is, Can it be defended? Of course this approach would seem to imply that not only children but also fetuses and embryos should count as persons in the strict sense, because an embryo is, on many accounts at least of the metaphysics of the situation, numerically identical with the mature human being it might become. Some proponents of this approach welcome that consequence, while others resist it.

But the entire "unity of the person" view, at least in the context of Kantian ethics, faces a fundamental difficulty: On Kantian grounds, no being can be considered a person in the strict sense at all unless it is at some stage of its existence a fully rational and morally responsible being (because rational and responsible agency are what its personality consists in). Yet not all children (much less all embryos and fetuses) ever reach that stage, and so it follows that they never do in fact become persons. Therefore on the "unity of the person" view, there seems no justification, on Kantian grounds, for saying that they are *ever* persons (at any stage of their history). This view seems committed to saying that a child who dies before it reaches maturity *never was a person.* We are required to say in retrospect that such a child *never had the status or rights of a person.*

Worse yet, it follows that we can never know about a newborn infant (not to mention a fetus or an embryo) whether it ever will be a person; its moral status seems, on this view, to be shrouded in uncertainty, or at best only presumptive rather than actual. Still worse than that, it follows on the "unity of the person" view that one could prevent an embryo, fetus, or even an immature child from ever becoming a person at all simply by killing it before it ever achieves personhood in the strict sense. And there would seem to be no possible moral objection to doing that, because the being in question never was and never will be a person. No proponent of the "unity of the person" view would welcome *that* conclusion.

Persons in the strict and the extended sense. I think Kantian ethics must therefore reject the "unity of the person" account. A more consistent Kantian approach is based on the idea that we can treat, or fail to treat, rational nature as an end in itself not only in the person of a rational being in the strict sense

but also in the way we treat other beings who are not persons in the strict sense. For instance, it would surely show disrespect for rational nature not to further its development to maturity in a child in whom it has already begun to develop. The same is true if we did not care about the recovery of rational nature by an adult who has temporarily ceased to be a person in the strict sense because of injury, disease, or some other incapacitation.

Thus in order properly to respect rational nature, we are required to treat some beings who are not persons in the strict sense in certain respects just exactly as if they were persons in the strict sense. Or, to put it another way, we are required to accord, at least for certain purposes, a status equivalent to personhood to some beings that simply are not persons in the strict sense. For instance, we should treat small children as having a right not to be killed, to have their well-being looked after, and their development toward maturity cared for. I propose that we apply the term *persons in the extended sense* to beings that are not persons in the strict sense but that should be granted a moral status (in the relevant respects) exactly like that of beings that are persons in the strict sense.

Persons in the extended sense do not have precisely the same moral status as persons in the strict sense. But they do not have a *lesser* status. If they lack the rational capacities to direct their own lives without guidance from others, then they cannot have the same right to direct their lives that persons in the strict sense have. We are permitted (even required) to behave paternalistically toward them, as we are not toward persons in the strict sense (MS 6:454). For the same reason, they are not held responsible for their actions in the ways that persons in the strict sense are.

Here we must also face up to the fact that who counts as a person in the extended sense is something that must be determined by those of us who are persons in the strict sense. For it is only persons in the strict sense who have the capacity to decide such questions, and they also bear the full responsibility for deciding them. Of course, that determination must be made for good and objective reasons. Persons in the strict sense may not simply satisfy their own desires or promote their own interests when there are good grounds for doing otherwise. Thus persons in the extended sense – once we have determined how far this status should extend – have just the same right not to be killed as persons in the strict sense, and we have the same obligations to consider their interests and treat them as ends in themselves that we have toward persons in the strict sense. In fact, precisely because they are *not* fully rational and self-governing beings that are competent to look after their own welfare, that welfare arguably has claims on our concern that should sometimes take priority over the welfare of beings that are persons in the strict sense – though not a claim that could encroach on the right of persons in the strict sense freely to direct their own lives.

The limits of personhood. The obvious question at this point is: Exactly at what stages of human life should beings be regarded as persons in the

extended sense? The answer to this question, as I have framed it, depends on how far our conduct in treating the human beings in question expresses due respect for the dignity of rational nature and how far it falls short of this or violates the dignity of humanity. It is relevant to the right answer to such questions not only how we are acting toward rational nature in our treatment of human beings who are not persons in the strict sense, but also whether in our conduct we duly respect this value in those who are persons in the strict sense.

For example, consider the question of whether a fetus, like an infant, is to be regarded as a person in the extended sense. That question should turn not only on whether our conduct duly respects the value of the (still merely potential) personhood (in the strict sense) of the fetus, but also on whether it duly respects the dignity of actual persons in the strict sense – in particular, the dignity of the person in whose body the fetus is developing. If that person is forced to bear a child she does not want, or if her right to control the life processes going on in her body is coercively restricted by others (as by either forcing her to have an abortion or by denying her one), then their conduct expresses extreme disrespect for the right of rational nature in her person. Regarding the question of whether an embryo *in vitro* is a person in the extended sense, that should turn on whether, in order to treat it as a person, some woman would have to be coerced into having the embryo implanted in her uterus and then compelled to carry it to term. Clearly if she would, then the embryo should not be judged a person in the extended sense.

I conclude that if granting to embryos or fetuses the same "right to life" that is thought to belong to persons in the extended sense would involve such coercive or invasive conduct, then it would constitute gross disrespect to rational nature to grant them that status. (Part of my reason for giving this answer is social and historical, having to do with the way human cultures have traditionally treated women, and how we should be trying to treat them now. I will return to this issue in Chapter 13, §2.)

The Kantian position, as I interpret it, should be that there is certainly some value in the potential personhood of an embryo or a fetus – or, I would equally say, of an unfertilized ovum, though not necessarily exactly the same value as that of an embryo or fetus. None of these entities has the sort of value that pertains to persons in the extended sense (carrying with it, for example, a coercively enforceable right not to be killed or destroyed). Generally speaking, and subject to modification in borderline or problem cases – the dividing line between a person in the extended sense and a nonperson whose life still has some value should be drawn at birth. The reason is this: Prior to birth there is no way of granting the status of personhood in the extended sense to the being without violating the right of persons in the strict sense, whereas after birth there is at least in principle the possibility of doing this.[9]

Kant's own statements. We might wonder what Kant's own position is on the personhood of a fetus or embryo. Of course he never distinguishes, as I have, between persons in the strict sense and persons in the extended sense. In fact he has all too little to say about such issues at all. Kant asserts, without any explicit argument, that even small children are persons, even that "the offspring is a person," and the parental duty to care for the offspring "follows from conception [or procreation] (*aus der Zeugung*)" (MS 6:280).[10] Kant declares in the same passage that it is "impossible to form a concept of the production of a being endowed with freedom through a physical operation" (MS 6:280). It appears he does not think the issue of when personhood begins can be settled directly by empirical inquiry. He even denies that the question of when personhood begins is properly conceptualizable empirically in biological terms.[11] But Kant does claim that "it is a necessary idea to regard the act of conception [or procreation] as one by which we have brought a person into the world without his consent and on our own initiative, for which deed the parents incur an obligation to make their child content with his condition as far as they can" (MS 6:280).

Of course it is one thing to say that parents should be thought of as bringing a person into being, and even that they have duties of care to their offspring from conception. It is quite a different thing to say that the offspring is a *person* from conception onward. The first two things Kant does appear to say; the third is something he never quite says.

There are passages in which Kant might be thought to be addressing issues about the personhood of fetuses or embryos at least indirectly. In the course of his condemnation of suicide as violation of a duty to oneself, he asserts that when a pregnant woman commits suicide, she is guilty of murdering her unborn child as well as of murdering herself (MS 6:422). On the other hand, in the course of discussing whether an unmarried mother who kills her infant to avoid dishonor is guilty of murder, Kant offers the argument (it is not clear how far he endorses it) that because an illegitimate child has come into being under conditions other than those recognized by the state, the state is not required to acknowledge its existence, or, therefore, to regard the causing of its death as the violation of anyone's rights (MS 6:336). The first passage seems to treat a fetus as a person, and by implication, abortion as murder; the second seems to claim that infanticide (much less abortion) is not murder when committed by an unwed mother to preserve her honor.

In both passages, however, the argument is being driven by quite another agenda than the question of the personhood of a fetus. In the first passage, Kant is concerned to press a controversial point that is part of his general theory of the imputability of consequences, namely that someone who does wrong is responsible for all the bad consequences of his wrongdoing, whether foreseeable or not. So he wants to heap up bad consequences as much as he can. In the second passage, he is trying to diagnose a conundrum that arises when unenlightened social attitudes – here, attitudes toward

unwed motherhood – force an individual to choose between taking a life and totally sacrificing honor; he wants to explain why we are reluctant to treat such homicides simply as cases of murder. In both cases, the local agenda probably drives him to assume for the moment certain views about the personhood of the unborn that he may or may not actually embrace. For this reason, I do not think that any determinate view on this issue can be reliably ascribed to Kant, even on the basis of the passages in which he seems to be discussing it at least indirectly. Even if Kant had stated an unambiguous position on the personhood of embryos or fetuses, the task of Kantian ethics would not be to follow his errors blindly but rather to consider what Kantian principles really imply and to interpret them correctly.

Some who advocate the "unity of the person" approach, as well as some non-Kantians, have expressed discomfort with the distinction between persons in the strict sense and persons in the extended sense. If they fear that the distinction involves granting a lower moral status to persons in the extended sense than to persons in the strict sense, then the response is that the whole point of the approach is precisely *not* to do that. Persons in the extended sense have the same *dignity* as persons in the strict sense.

What often really bothers them, however, is the whole thought that moral status or personhood is tied to some property other than membership in the human species, some property that not all members of this species possess. As I see it, however, that thought is simply the inevitable result of requiring a *reason* for granting human beings the status of persons. For it is self-evident that membership in some biological species can never by itself constitute such a reason. (Peter Singer's objections to "speciesism" are obviously correct, at least to that extent.) To hold that we should regard all humans as persons because they are members of *our* species seems no better than regarding as persons only those who share our nationality or religion or skin color. To argue that certain entities are persons because they are members of a rational species, when they are not in fact rational beings, makes no more sense than arguing that children are already human adults because they belong to a species whose mature members are human adults. The right reply to such an argument is: You simply can't get there from here, at least not using that road.

If, however, we have a good reason for holding that human beings should have the moral status of persons, then that reason will have to consist in some property other than their simply being a member of our species, and whatever property it is (whether rational nature or something else), some members of the human species will possibly not have it. Thus those who do have it will be persons in the strict sense, and questions will inevitably arise about whether, why, and in what respects we should grant a like status to humans who do not have that property (thus invoking some concept like personhood in the extended sense). I conclude that discomfort with the distinction between persons in the strict sense and persons in the extended

sense must display either a stubborn refusal even to ask for a reason why human beings should have the moral status of persons or else the unthinking acceptance of a patently unsatisfactory reason.

5. The Moral Status of Nonrational Animals

Kant supposed (as most of us still do) that human beings are the only creatures on earth that have the capacity to set ends according to reason, devise means to them, form a conception of their own general well-being or happiness, and regard themselves as legislators as well as subjects of moral laws. But he thought there were probably finite rational beings on other planets (ANG 1:349–68, VA 7:331), and it is still possible that we may find such beings (though the likelihood now seems far lower than it pleased people to think it was even a generation ago). Some people now think that the mental capacities of the higher primates or other mammals essentially qualify them as rational beings. If empirical research were to support such claims, then Kantian ethics should be the first to accept them and modify accordingly the system of moral rules and duties. There would still be the same questions, however, about how to treat other living beings who clearly do not share in rational nature, but do share in life, sentience, purposiveness, caring, or other substructures, fragments, and analogues of rational nature.

The approach taken here to the way Kantian ethics should regard the treatment of nonrational living things is along the same lines as above: We should ask whether our conduct toward them shows due respect for the dignity of rational nature. Yet it might seem *prima facie* as though our conduct toward beings that are not and never could be persons in the strict sense could show neither respect nor disrespect for the dignity of the rational nature – and therefore that this Kantian approach could say exactly nothing about how we would treat nonrational living things. But I claim this is not so. For the life of many nonrational living things actually shares many features with that of persons in the strict sense; and the way we treat that life, regarding these features, can and must be interpreted as expressing either respect or contempt for rational nature. I do not think that any nonrational living thing should ever be accorded the status of a person in the extended sense. Yet Kantian morality does, it seems to me, forbid certain kinds of conduct toward such beings and justifies, or even requires, some positive concern for their welfare.

"Can they suffer?" Jeremy Bentham famously said, regarding nonhuman animals: "The question is not, Can they reason? Nor Can they talk? But, Can they suffer?"[12] This is usually understood as a rejection of the Kantian position on such questions. For as a hedonist, Bentham apparently bases moral status not on the dignity of rational nature but rather solely on the capacity to feel pleasure and pain. And this is clearly different from the Kantian position. Yet I claim that Bentham's idea here is in general terms

not inconsistent with Kantian ethics but is instead a corollary of the Kantian position. I would even claim that Kantian ethics provides a better justification for it than Bentham's hedonism – a shallow empiricist doctrine that cannot account properly even for the values it assigns to pleasure and pain in human beings.[13]

The happiness of human beings is an end of morality because setting this end shows respect for rational nature. Nonhuman animals, like human persons both in the strict and the extended sense, have desires, preferences, and a capacity for pleasure and pain. In these respects, they are like human beings, and their desires, preferences, and sensibilities are even analogous to the rational capacities of humans in the way they direct the behavior of the animals. Therefore, all other things being equal, when we frustrate the desires or preferences of persons (both in the strict and extended sense) or when we cause them suffering or fail to promote their welfare, we show disrespect for their rational nature (actual or potential). Likewise, when we wantonly or maliciously frustrate the desires and preferences of nonrational animals, or cause them pain, we act in a manner analogous to the way we act when we show disrespect for the rational nature of persons. The capacities of animals can be said to belong to rational nature as parts, or necessary conditions, or as its infrastructure, so to speak. Our conduct toward these animals can therefore be approved or condemned by Kantian ethics based on what it expresses toward the value of rational nature, even though nonhuman animals never actually possess in themselves the full capacities of rational nature that make a being into a person in the strict sense.

Nonhuman animals do not have the capacity to reason or to talk. Therefore, beyond making the obvious point that they are not persons in the strict sense, whether they have or lack these capacities is irrelevant to how we should treat them. Bentham is therefore correct in telling us not to ask about these matters when we are deciding how to treat animals. What is relevant, because it relates their capacities to those of rational nature, is the fact that they can suffer, and desire, and sometimes also care – about members of their own species, or even occasionally about members of other species, such as humans. Bentham is therefore also correct in telling us what we should ask about these capacities, for they are the relevant ones. Bentham is correct, however, not because Kant is wrong, but because Kant is right.

It would be a gross misunderstanding to take what I have been saying as the endorsement of some principle to the effect that we must place value (or equal value) on anything and everything that could be regarded as a part, or necessary condition, infrastructural element, or analogue of rational nature.[14] There are surely many such things – some physical or chemical prerequisites of life, for example – that we can obviously treat as having little or no value without showing any disrespect for rational nature. Some animals possess the capacity to care (about their young, about other members of their species, or even about human beings). This capacity is clearly a larger and

much more immediate component of rational nature than the mere capacity to show a preference for moving in one direction rather than in another (as an insect does) or even the capacity to feel pleasure and pain. Hence, from a Kantian standpoint there is reason to be concerned more about animals that are capable of caring about others than about animals that are not. The relevant judgments here are hermeneutical or interpretive in nature. Such judgments are notoriously not derivable by strict deductions from general principles. Rather, the right way to look at the matter is that for Kantian ethics the way we arrive at conclusions about determinate moral rules, for the treatment of persons or anything else, is that these are the results of interpreting the supreme principle of morality – here, in the form of FH. The capacity to care about others is also a *human* capacity that belongs to human beings (such as small children, or Alzheimer's patients) who do not have the full capacities of rational action that make normal human adults persons in the strict sense.[15]

There are important factual questions that seem also to involve questions of interpretation, concerning to what degree some of the higher mammals – chimpanzees, for example, or dolphins – share in fragments of the capacity we conceptualize as human rationality. It is controversial, for instance, whether chimpanzees have a sense of "self," and to what extent the sense of themselves that they may have involves participation in the capacities for which we should value human beings. It is an interpretive judgment that we should protect the external freedom of a person in the strict sense to govern his or her life; that we should treat children, but not fetuses or embryos, as persons in the extended sense – that is, beings with the same right to life and concern for their welfare as persons in the strict sense; and that we should treat nonhuman living things – the higher mammals, for instance – as beings whose health, desires, and contentment matter to us, even though they are not persons in either the strict or the extended sense.

Kant on the treatment of animals. What I have just said about the treatment of nonhuman animals appears to differ significantly from what Kant himself says in the *Metaphysics of Morals* and elsewhere. Kant thinks the notion that we have duties to nonhuman animals results from an "amphiboly in moral concepts of reflection," where a duty to ourselves appears to us as a duty to beings other than ourselves (MS 6:442). Kant thinks that we have a duty to ourselves to display the moral perfection of kindness toward the suffering of sentient beings and avoid the moral imperfection of callousness or cruelty to them. Not fulfilling these duties, he argues, will corrupt our dispositions toward other people and their happiness or suffering. This doesn't look like the account I have just given. But the differences between my account and Kant's are not as great as they may at first appear.

It is important that Kant does *not* treat our duty regarding animals as a duty to others – as though he thought we should develop habits of kindness rather than cruelty toward animals merely for the benefit of the humans toward

which we can be expected subsequently to display these traits. Rather, he regards the virtues of kindness and gratitude toward animals as in themselves perfections of our character and the vices of callousness and cruelty toward them as in themselves failings or imperfections. As I understand Kantian principles, this means that kindness toward animals itself complies with duties whose principle is treating rational nature as an end in itself, and callousness and cruelty violate duties based on the same principle. For this reason, the arguments he uses presuppose something like the account I have derived from Kantian principles.

There are also remarks in Kant's lectures that suggest something like this account. Animals, he says, are "analogues of humanity," and this is why we have duties that are also "analogues" to our duties to human beings. "If a dog, for example, has served his master long and faithfully, that is an analogue of merit, hence I must reward it, and once the dog can serve no longer, I must look after him to the end" (VE 27:459; cf. VE 27:710).

> If the acts of animals arise out of the same *principium* from which human actions spring, and the animal actions are analogues of this, we have duties to animals in that we thereby promote the cause of humanity . . . The more we devote ourselves to observing animals and their behavior, the more we love them, on seeing how greatly they care for their young; in such a context, we cannot even contemplate cruelty to a wolf. (VE 27:459)

Kant's use of the term "analogy" here also suggests his theory of analogical or symbolic language, in which words are used that do not signify something directly but indicate it indirectly by employing a procedure of the understanding that is like that through which it might be directly signified (P 4:356–60; VpR 28:1023; KU 5:351–4). We have duties in regard to animals because their behavior and their needs have, in relation to the worth of rational nature, a significance that is similar in certain respects to the behavior and needs of rational beings. Consequently, we have duties of a comparable significance, based not directly on their animal nature but on the worth of humanity, to which the animals, their life processes, and behavior are "analogues."

Kant's own more specific views about how animals are to be treated seem to me generally sensible and decent (if not particularly remarkable). Kant thinks it is permissible to kill animals for human ends (such as for food); but he insists that this should be done as quickly and painlessly as possible (MS 6:443; VE 27:459–60). He regards killing animals for mere sport as morally wrong (VE 27:460). He insists that domestic or work animals should not be overworked, and that an animal, such as a horse or dog, that has served us well should not be cast aside like a worn-out tool when it is too old to perform its task but should be treated with gratitude and affection, like a (human) member of the household, and be allowed to live out its days in comfort. Kant regards as morally abominable "agonizing physical experiments [on

animals, carried out] for the sake of mere speculation, or whose end can be achieved in other ways" (MS 6:443). He praises Leibniz for taking the trouble to place a worm back on its leaf after examining it under a microscope (KpV 5:160; cf. VE 27:459).

These seem to me generally the right kinds of conclusions to draw, based on Kantian principles. Yet some Kantians I know are vegetarians – on what they regard as Kantian grounds. Other Kantians I know think Kantian principles require even that embryos and fetuses be regarded as persons with a coercively enforceable right to life. The moral I draw from all this is that fundamental Kantian principles do not, all by themselves, necessarily determine in advance the answers to these moral questions. It is, as I have already argued, a profound misconception of moral philosophy – of its proper structure and the role of a fundamental principle of morality in it, that the fundamental principle all by itself should attempt to do anything of the kind. What is important is that the fundamental principle of morality should correctly orient our values in thinking about moral questions. The task of reaching, or even advocating, determinate answers to them belongs to a separate stage of moral philosophy, the stage that interprets the fundamental principle and derives from it moral rules and duties. What I have been arguing in the last two sections in fact more properly belongs to that stage of moral philosophy, which I will take up for itself only later, in Chapter 9.

13. It might also be thought to agree with the following claim: "One must be able to will that the maxim of our action should become a universal law: this is the canon of the moral judgment of this action in general" (G 4:424). This seems to say of FUL (or perhaps FLN) that it is the canon (or standard) of moral judgment – which is what Kant says about the "universal formula" at G 4:436. But it seems highly questionable to take this statement to be about FUL or FLN in particular, as distinct from other formulations of the moral principle, at a point in the exposition at which no other formulas have yet been introduced. It is similar in regard to Kant's statement a bit earlier that "The categorical imperative is thus only a single one, and it is this; . . . " followed by a statement of FUL (G 4:421). At a point where Kant has no other formulas with which to compare FUL or FLN, it seems a mistake to read him as speaking about these formulas as distinct from others yet to be introduced. It seems more reasonable to take both statements to be general ones about the principle of morality Kant is in the process of deriving. Which formula of this principle is the definitive one, or the one best suited as a canon of moral judgment, is at the time of these statements as yet undetermined. Later stages of the exposition clearly indicate that the definitive formula, and the canon of moral judgment, is FA rather than FUL or FLN.

Chapter 5. Humanity

1. It follows directly that metaethical antirealists can respect anything at all only by acting in a way that implicitly conflicts with their own doctrines.
2. Peter Singer is no Kantian, but I think a Kantian could (or even should) agree with a lot of what he has said against the doctrine of the sacredness of human life. See Singer, *Rethinking Life and Death* (New York: St. Martin's Griffin, 1996).
3. She first did so in the article "The Formula of Humanity," first published in *Kant-Studien* 77 (1986), later in Korsgaard CKE, p. 128. The idea is key to her subsequent writings, especially Korsgaard SN. Despite my disagreement with her on this point, my own interpretation of Kant's argument for FH obviously has some strong affinities to hers, as critics of both of us have frequently noted. But there are significant differences between us, which the critics have often failed to appreciate. I hope the present discussion may make some of this clearer.
4. For instance, in Korsgaard FC, pp. 92–3.
5. Kant also says here that this involves attributing to other rational beings the same property we attribute to ourselves: "It is not enough," he says, "to ascribe freedom to our own will, on whatever grounds, if we do not have sufficient grounds to attribute the same quality also to all rational beings" (G 4:447). And he concludes his argument by claiming that because any rational being must "act under the idea of freedom," it must extend the same thought to all: "It must regard itself as the author of its principles independently of alien influences: consequently it must, as practical reason or as the will of a rational being, be regarded by itself as free: i.e. the will of a rational being can be a will of its own only under the idea of freedom and must therefore with a practical aim be attributed to all rational beings: (G 4:448). We will return to this point in Chapter 7. For now, however, it should be clear that Kant is not in agreement

with those approaches that think we are to regard ourselves as free only from the "standpoint of the agent" or the "first person"; this would be insufficient to establish what his argument requires. For such approaches see Korsgaard, CKE, pp x–xii, and Hilary Bok, *Freedom and Responsibility* (Princeton, N.J.: Princeton University Press, 1998).

6. We may sometimes have to sacrifice some human being's interests, or even his life, but as we have seen, this is not necessarily the same as sacrificing the *dignity* of the person. There are several different senses in which Kant speaks of something as having "absolute" value. This is one of them. Another is that what has absolute value has it irrespective of any relations to anything else (in particular, irrespective of any relations to the way it is considered or judged – G 4:439). Yet another is the goodness without limitation that belongs to every good will (G 4:394). Kant also speaks of a certain kind of good will, one that is not only good but an "absolutely good will," when it does not merely act on good maxims but acts on the *principle* of acting only on maxims that are universally legislative (G 4:437). This plurality of senses of "absolute" may at times be confusing, but we need to be aware of it if we are not to make mistakes (or attribute to Kant mistakes that he does not make).
7. The one exception might be R 6:26n, where Kant says that from the fact that a being has reason it does not follow that it is capable of representing its maxims as universal laws or that its reason is capable of being unconditionally practical. If Kant's point here is merely that the concepts of humanity and personality are different, then what he says is correct. But if he is claiming that a being might have humanity without personality, then I think he is wrong, for reasons given in the text.
8. See Ludger Honnefelder, *Die Einheit des Menschen* (Paderborn: Schöningh, 1994). Christine Korsgaard has defended a similar view to me in conversation, though I have never seen her do it in print and cannot be sure I am understanding her correctly. But what I recall her saying, putting the matter in Kantian terms, is that we apply the same concept of noumenal personality to a newborn child as we do to the mature human being it will become. Honnefelder welcomes the possible consequence of such a view that it extends personality to embryos and fetuses, while I take Korsgaard to resist this consequence.
9. The issues we are discussing here are, in Kantian terms, issues of right, not of ethics. They concern duties and claims that may be coercively enforced. It is a separate question whether the value of a fetus, and of its development to the point of birth, might constitute a reason for a woman to regard as an ethical duty her carrying it to term even at considerable cost to her own health and welfare. Probably there is such an ethical duty, at least in many cases. But those who would deny a woman even the right to make a choice whether to comply with this duty have thereby utterly forfeited their standing to argue with them about such issues. Someone who would deny you the freedom to make a choice that is rightfully yours to make has no business trying to tell you how you ought to make it. The effrontery of those who picket abortion clinics should therefore be met with stony stares of contempt by the women in whose lives they are trying to meddle.
10. The word *Zeugung* can mean 'conception' or 'begetting' but also something more general, such as 'generation' or 'procreation'; hence it does not

definitively settle either way whether Kant thinks the fetus is a person prior to its birth. I do not deny that it might reasonably be interpreted to favor an affirmative answer to that question, but if Kant had intended to favor such a view, he certainly could (or even should) have expressed himself more explicitly. In any case, he never *argues* for this stronger view.

11. In a recent paper, Patrick Kain employs some of Kant's views about biology to argue that he thinks of personhood as belonging to every human being, beginning with conception, Kain, "Kant's Defense of Human Moral Status," *Journal of the History of Philosophy* (forthcoming). This would be an alternative route to something like the "unity of the person" view, and, if accepted, it would yield the conclusion that every member of the human species, from conception, is a person as regards its rights and moral status, even if this being never develops the capacities necessary to be a person in the strict sense. No doubt such a conclusion would be welcome to many people today, especially those on the conservative side of issues about the permissibility of abortion, embryonic stem-cell research, and so forth. Kain's reconstruction of Kant's position is informed and in some ways plausible, though it seems to me more a speculative reconstruction than a compelling account of what Kant actually thought. And the attempt to bring biology to bear on it seems to fly in the face of Kant's declaration that we can form no proper concept of the beginning in time of a free being. Even if what Kain formulates were Kant's actual view, however, this view would depend heavily on Kant's controversial version of what is for us a clearly outdated biology – and even involves a willingness on Kant's part to entertain the theory of "preformation" regarding human persons – a view he otherwise consistently rejects regarding the natural genesis of living things. The conclusion to be drawn from Kain's account for present-day Kantian ethics is that if this is Kant's own account of personhood and moral status, then it is no longer a viable option, and Kantian ethics should seek a different one that can be grounded on Kantian principles that might still be tenable. That, in effect, is what I am doing in this section.
12. Jeremy Bentham, *Principles of Morals and Legislation* (New York: Hafner, 1948), p. 311n.
13. It is not easy to say how, or how much, different living things do, or can, feel pleasure or pain. I have spoken to biologists who say that the nervous systems of most invertebrates are too primitive for us to ascribe much real pain ("in our sense") to them, but also to other biologists who have rejected such a judgment with horror. It is common for chefs to say that a lobster does not suffer when plunged into boiling water, but the heroine of the German film *Bella Martha* (2001) (English title: "Mostly Martha"), who is a master chef and hardly a sentimental character, holds that this is terribly cruel and insists on a different method of killing a lobster before boiling it. There may be deep conceptual confusion at this point in our entire way of thinking about mental states, perhaps of people as well as animals.

    If mental states are real, then they must exist, be what they are, when known from a variety of perspectives, not a single perspective. "What can be known only from a single perspective" is almost a definition of the merely imaginary or unreal. So if our mental states are real, their reality is constituted in part by others' interpretations of them and justified attitudes toward them. But regarding

mental states, especially states such as pain, we tend to think that the only perspective on it that matters is that of the being experiencing it. This would seem to be a *moral* stance on our part, or perhaps a premoral stance that enters into morality at a fundamental level as one of its ingredients – that, at any rate, is what Rousseau thought (Rousseau D 37–8, Rousseau E, 41–2, 181–92). But it also seems that for rational beings, their valuation of the being whose states are in question plays a role in the attitude they take – a state of affairs that Rousseau acknowledges in his typically provocative manner by imagining a civilized man whose neighbor is being murdered right under his window: "He has merely to place his hands over his ears and argue with himself a little in order to prevent nature, which rebels within him, from identifying with the man being assassinated" (Rousseau, D, p. 38). We miss the point if we see nothing more in this than a condemnation of the civilized man. For Rousseau's point is precisely that he and his audience are all beings of exactly that kind, in whom natural sympathy has necessarily been transformed by society and reason. We are supposed to see in the absurdly cruel conduct of the civilized man not merely something to be avoided but the very predicament to which we are all in one way or another condemned by our social condition. In other words, if there is to be any hope for any of us, it will have to depend on the quality of the arguments we give ourselves while standing there with our hands over our ears.

As it seems to me, "sympathy" in this sense plays a role not in our formulation of fundamental moral principles but in our interpretation of them in moral rules and in the application of these rules. But when we are unsure what our stance on the pain of some being is, because it is a being whose nervous system is very different from ours, we would seem to have lost our moorings. The line separating callousness from excessive sentimentality seems extremely hard to draw. Those who think anything is obvious here, or that it can be settled empirically in any easy way, seem to me very much mistaken.

14. I am grateful to Allan Gibbard for bringing this issue to my attention.
15. See Agnieszka Jaworska, "Respecting the Margins of Agency: Alzheimer's Patients and the Capacity to Value." *Philosophy & Public Affairs* 28(2) (1999), pp. 105–38.

Chapter 6. Autonomy

1. See Friedrich Schlegel and Friedrich Schleiermacher, in F. Beiser (ed)., *The Early Political Writings of the German Romantics* (Cambridge: Cambridge University Press, 1996), pp. 132, 155, 174–5.
2. "It is not such a very long step from Kant to Nietzsche, and from Nietzsche to existentialism and the Anglo-Saxon ethical doctrines that in some way resemble it. In fact, Kant's man had already received a glorious incarnation earlier in the work of Milton. His proper name is Lucifer." Iris Murdoch, *The Sovereignty of Good* (London: Routledge, 1970), p. 80.
3. The term 'theonomy' is most associated with the theology of Paul Tillich, though it was not original with him. See Tillich, *Systematic Theology*, Volume III (Chicago: University of Chicago Press, 1963), pp. 249–74. Cf. Franz Rosenzweig, *The Star of Redemption*, tr. W. W. Hallo (Notre Dame, Ind.: University of Notre Dame Press, 1985), pp. 28–30, 66–9, 112–16, 176–82. There is an interesting