

A JUSTIFICATION OF MORAL VALUES

By Arnold Zuboff

This paper combines two published pieces.

The first section is a slightly modified version of “Why Should I Care about Morality?”, Philosophy Now, Issue 31 (2001), pp. 24-27.

The rest of this paper is a slightly modified version of “Morality as What One Really Desires”, Midwest Studies in Philosophy, vol. XX, (1995), pp. 142-164.

Why Should We Care About Morality?

Let's inquire into the nature of morality--and, more particularly, into the authority that it seems to have in the judgments of most of us. I think a certain story can help us in raising the question of where it gets that authority.

Imagine that someone we shall call Gyges, after a character similarly used by Plato in a basically similar story, is seated at a table. Just before him on the table is a small console with a single button on it. Let's say he knows that if he pushed that button a distant stranger, who would otherwise be fine, would be killed. Gyges also knows that if he pushed that lethal button, he, Gyges, would be given £10 that he otherwise would not have. We are going to look into whether Gyges has any reason based purely on morality not to push the button.

It is vital that we rule out of our story, if it is to be useful to our questioning of the authority of morality, any possibility that Gyges be punished if he pushes the button or that he in some way be rewarded if he doesn't push it. For if we give him the fear of punishment or the hope of reward as reasons not to push the button, we have not then clearly isolated whether he has a reason not to push the button in its being morally wrong to do so. We are wanting to know whether morality in itself has an authority here for him, but his own punishment or reward carries only the authority of the sort of obvious self-interest that is often distinguished from moral motivation. Therefore we shall say something like Gyges can be sure that the death he might choose to cause would have the perfect appearance of an accident having nothing to do with Gyges. So Gyges would be perfectly safe. Let us add that the remoteness of the stranger insures that there would be no other possibilities of personal loss or gain for Gyges in either the stranger's death or his continued life.

The question, then, is this: Gyges has a slight but undeniable reason to push the button, the self-interested reason that he will by doing so acquire £10 that he otherwise would not have. But does he have a reason not to push the button?

Let me now mention that I have often had occasion to pose this question to people with some interest in philosophy, since I have often used this problem as a topic for discussion in interviews with applicants to our department. In these interviews, after I have described the situation in which Gyges finds himself, I add that, although he perhaps sounds a bit nasty, in that he is sitting there considering whether to kill someone for £10, Gyges can at least be credited with being open-minded: and he will always be genuinely interested in any advice he may be given about whether it makes sense for him to behave one way or another. And, I continue, my interviewee is now to have an opportunity to offer to Gyges (for whom I will be speaking) any such advice that seems appropriate.

In a small number of cases the interviewee will say at the start that Gyges only has reason to press the button and collect his £10. In a still small but larger number of cases the

interviewee will arrive at this conclusion after some attempts to come up with a reason for Gyges to refrain from the killing. But most people will first make attempts to explain how Gyges has a reason not to push the button and then, while more or less confidently retaining the belief that there is such a reason, will feel forced to quit trying to find it. These people, and some of the others, often seem not just surprised but also somewhat relieved when I say I can offer my own answer to the challenge. And I shall do so here after I have first considered the sorts of attempts people make and my criticisms of them.

These criticisms take two general forms. Sometimes I argue that an attempted solution amounts to merely reintroducing the problem in other terms. Perhaps the simplest example of this is when the interviewee points out that the killing of the remote stranger is morally wrong. In response to this I bring out that Gyges knows that pushing the button would be a sort of action that is regarded as morally wrong. Let's say he himself regards it as "morally wrong". But what he wants to know is whether this description of the action carries with it for him a reason for not performing the action. If it does, that reason has yet to be explained. Similarly, if the interviewee has said something about violating rights or the sanctity of life, I say on behalf of Gyges that he is extremely interested in these as possible sources of a reason for him not to push the button, but he still needs to have explained to him not only what such alleged rights or sanctity are, what they consist in, but how they are supposed to compete with the straightforward self-interested motivation of £10 by giving him a reason not to push the button. So these are just ways of raising the main question again without yet answering it.

Some other ways of merely reintroducing the question inspire a response from Gyges that illuminates further the character of the challenge. For example, the interviewee might simply claim that Gyges *should* refrain from pushing the button. This demand leads Gyges to distinguish between moral and non-moral uses of the word 'should'. Gyges says he understands very well how the advice that he *should* take up skiing is pointing out that he has some sort of reason to do so. The idea would be that Gyges himself would benefit from skiing. That gives him his reason. This use of 'should', however, has no moral significance and carries none of the supposed authority of morality. But the command (rather, it seems, than advice) contained in saying that Gyges *should* refrain from pushing the button bids him to sacrifice a benefit, the £10, in order that a remote stranger not suffer a loss. Why *should* Gyges do that? How could he have a reason to do it?

Also illuminating is his reply to the following attempt to give him a reason not to push the button: The interviewee asks Gyges to put himself imaginatively into the shoes of the remote stranger he might kill. "How would *you* like it if someone pushed such a button on you?" Gyges replies that he'd *hate* to have someone do that to *him*. But that's just the point. In *this* case it would be somebody *else*, *not* Gyges, who died and it would be Gyges, not somebody else, who received the £10. Of course, Gyges would have every reason not to want someone else to push the button on Gyges. But how can that fact possibly give Gyges any kind of reason not to push the button on somebody else? He can put this in a somewhat different way. It is obvious why it is *bad* for the remote stranger if Gyges pushes that button. The stranger loses his life. But how does this being bad for that stranger somehow (perhaps magically?) spread itself across to it being *bad* also for Gyges if he pushes the button? Gyges understands the force in the non-moral uses of 'good' and 'bad', but not yet the moral use.

An interviewee will sometimes say that the enormous value to the stranger that the rest of his life is likely to have, and perhaps also the value of his continuing to live for the stranger's family and friends, can be presented to Gyges as giving him a reason not to push the button--especially when in pushing it Gyges would be gaining only the relatively microscopic value of £10. Gyges replies to this that he completely agrees that the value of

the stranger's life for the stranger would be likely by far to outweigh the value to Gyges of getting £10. Yet, Gyges asks, how is such value for the stranger to be counted as any value at all for him, for Gyges? The value of the £10 isn't much, but its value would be value for Gyges. And, if the enormous value of the stranger's life cannot be counted as any value at all for Gyges, then how can that value in itself give Gyges any reason to be acting with regard to it?

Often an interviewee will make a point in some ways related to the one about putting oneself in the other's shoes: Is it not a good thing for Gyges himself that at least many others in his society are not as dismissive of morality as he is considering being, even in cases where they could get away with being immoral? How would Gyges like it if he lived in a society of self-centred schemers? Gyges replies that he would hate living in such a society. He is very happy if other people are somehow so respectful of morality that they would not push the button on him. Moreover, he sees the advantage to himself in making others think that he too subscribes to morality; they'll tend to treat him better if they think that. But how does any of this give him a reason not to push that button when he can do so with perfect impunity? Since the killing will seem like an accident, there is not even the extravagant worry that it would somehow add to a general undermining of the moral values he is happy for others in his society to hold. What he does regarding the button is simply distinct from any question of how society will treat him. Perhaps one wants to try the point that the morally good behaviour of others makes it unfair for Gyges to ignore morality. One might want to claim that it places him under an obligation. But this just raises the questions, how does otherwise being unfair or betraying an obligation give Gyges a reason for not pushing the button and giving up £10?

So far I've criticised attempts to provide such a reason for Gyges on the ground that instead of meeting the challenge they merely encourage Gyges to ask his question about the authority of morality in somewhat varying terms. But there is a second kind of criticism that I often find myself making. It applies already to one element in the last-discussed attempt, the anyway mistaken notion that Gyges somehow is in danger of losing the benefit of living in a society in which others wouldn't push the button on him. My criticism is that the interviewee is smuggling into what is supposed to be a reason for not pushing the button that is based purely on morality a threat of punishment or a prospect of reward, the sort of obviously self-interested motivation that we have been trying to eliminate from our story in order to see what authority morality in itself may possess. Sometimes an interviewee points out that such killing as Gyges is contemplating seems to be against the will and commandments of God. (Occasionally God is brought in as standing behind rights, obligations or the sanctity of life.) My response is that the involvement of God seems to me to take two possible forms. It may be that God is thought to give Gyges a reason not to push the button because to do so would bring down punishment from God and count against receiving God's rewards for the good. But this is precisely the sort of self-interested motivation that distracts us from the question of morality's own authority. This kind of calculation has no more to do with the authority of morality than would Gyges being worried that the devil might punish him for *not* pressing the button if he thought that the devil was more powerful or more interested than God. So we must eliminate this distraction by adding to the story that Gyges has a special deal with God, who will be turning a blind eye to whether Gyges presses the button. But the other form that God's involvement might take is that one might think that there is a reason for obeying God in that God has a perfect understanding of morality. But that point just returns us to my criticism that we have not yet explained how morality itself gives Gyges any reason not to push the button.

A very common attempt to give Gyges a reason not to push the button is an appeal to his conscience. How can he live with himself if he takes a life? My response to this has the

same double character as my response to invoking the will of God. The pain of conscience might be meant here as nothing more than an internal punishment that Gyges may inflict on himself. Fear of that is not moral motivation. It may be that I would love to eat a luscious dessert but I know that if I do I will suffer from indigestion. That would give me a reason to refrain from eating the dessert. But if I had a pill that would fend off indigestion I would no longer have that reason for not eating the dessert. So we need something like a pill that Gyges could take to forestall the pain of a bad conscience if we can regard that as nothing more than an internal punishment that distracts us from considering purely moral motivation. And I have just the pill we need. There is a hypnotist standing by who can hypnotise Gyges into forgetting perfectly that he pushed the button. The result could be Gyges finding £10 in his pocket and not remembering, ever, how it got there. Or Gyges could have the hypnotist make him believe, after he has pressed the button and done the killing, that by pressing the button he had saved someone's life and then also received the £10 as a reward. So Gyges could end up with £10 and a warm glow of virtue, while, as would then be unknown to him, the remote stranger is dead. What reason could Gyges have for passing up that? On the other hand, conscience might have been invoked with a deeper significance. It might be thought that Gyges somehow has a reason not to do away with a properly informed conscience because his conscience connects him with morality and morality in itself must have importance for him. But such an invoking of conscience in a reason not to push the button depends, of course, on the very understanding of the authority of morality that has so far eluded us. So this deeper involvement of conscience falls into the category of merely reintroducing the challenge.

But what if Gyges is the sort of person who tends to feel sympathy for others? That could give him a rather powerful reason for not pushing the button. Now, it may be, as Hume argues, that when I deal with others purely on the basis of a feeling of sympathy (also referred to by Hume as "disinterested benevolence"), I cannot be thought of as acting from what we usually call "self-interest". If, for example, I give my life for others out of sympathy for them, is that naturally described as acting from self-interest? Yet, as Hume himself stresses, it is merely an accident of my character whether, and to what extent, I possess such motivation. And that is why Kant, unlike Hume, rejects sympathy as accounting for the authority of morality. In the case of Gyges, for example, we want to say that he has a reason not to kill the stranger that is based purely on morality quite apart from whether he happens to feel sympathy for the stranger.

Let's approach this consideration of sympathy from another angle. Does Gyges have a reason to indulge any feelings of sympathy he may have when doing so would stand between him and acquiring £10? Recall that there is a hypnotist standing by who could hypnotise Gyges into forgetting he has acted against his sympathetic responses or, better yet, who could hypnotise him into thinking that he has indeed indulged his sympathetic feelings and also received £10 for doing so, though the stranger is dead. If sympathetic desires for the good of others are thought of as really aiming at nothing more than a feeling in the desirer that the desires have been gratified, then the option of the hypnotist would always be preferable to not pushing the button even for a strongly sympathetic Gyges.

But this conception of the aim of sympathy seems to me crucially incomplete. I have a strong desire that my brother, who lives on the other side of the Atlantic Ocean, be doing well in his life. I also would like to know that he is doing well--and knowing this would make me feel good. But it is a mistake to think that the real point of my wanting my brother to do well is that a consequent thinking he's doing well will give me a nice feeling. Imagine that I am asked to choose among the following possibilities (and that immediately after my choosing I will be hypnotised into forgetting that I was asked to make a choice): I can choose between, on the one hand, my brother doing well while I am guaranteed always to

believe falsely that he is doing badly, and, on the other hand, my brother doing badly while I am guaranteed always to believe falsely that he is doing well. Now, my feelings do have *some* value in these matters. By far the best situation would be my brother doing well and my believing that and feeling good about it. But from the less appealing choices I have been given, it makes perfect sense for me to choose my brother doing well while I suffer thinking he is doing badly. This is because my sympathetic desires don't have to be aimed at producing some gratification in me. They can be aimed directly at a situation outside me--in this case the situation of my brother. In fact, if the only point of my desiring good for my brother was that I come to feel the gratification of the desire (and that I escape a feeling of its frustration), that strikes me as degenerate.

So, if Gyges happens to be a sympathetic sort, he could have a reason not to push the button even if he'd not only lose the £10 but be hypnotised into thinking falsely, and painfully, that he *did* push the button. The reason could be simply the desire that the stranger not suffer harm. But, once again, that reason is dependent on whether Gyges happens to be sympathetic--and sympathetic in the non-degenerate way, in the way that aims directly at the good of the stranger. He might just as well be disinterestedly malicious in a way that aims directly at the harm of the stranger. And that could give him a reason beyond the £10 to push the button. So we have not yet arrived at a reason for not pushing the button that is based purely on morality. Our consideration of how desires can aim directly at objects outside of us, however, is, I would maintain, turning us at last in the right direction.

I'll begin *my* account of Gyges' reason not to kill the stranger by using another example. Imagine that I have before me on a table a cup containing a thick, brown, steaming liquid. I want to drink that stuff because I think it is hot chocolate. But it is really hot mud. Well, in that case I don't *really* desire to be drinking that stuff. And neither is it in my self-interest to do so.

This example brings out the way in which desires depend on beliefs. I only ever desire a thing because of what I believe it to be. And since beliefs are correctable, so are desires.

From this observation I arrive at a sweeping principle: My only *real* desires are those I would have if I had a perfect grasp of everything involved. If there is any desire I have only because my grasp of what's involved is less than perfect, then that cannot be among my real desires. And gratifying that desire cannot be in my real self-interest. The principle going along with this that governs my actions must tell me to act, *as far as possible*, as I would want myself to be acting with a perfect grasp of everything involved.

This perfect grasp that defines my real desires and my best course of action, what is it like? It would have to be like the all-penetrating knowledge that is often attributed to God. It would have to embrace not only the full experience, from behind the eyes (or other sensors), of every sentient being but also every potential development of experience. It would include within it, then, all the motivation of all of the various systems of desire, but it would also have the correction of all that motivation in light of the perfect grasp. The overall result must be a desire for the reconciliation of all systems of desire. And that, I would claim, is the concern that defines morality.

What I am saying, then, is that everyone's real self-interest merges together as what would be wanted in the single perfect grasp of everything. What Gyges really wants, whether he realises this or not, is to do what he would want to be doing if he were grasping the full value of the life of the stranger as well as £10. And even from the actual, limited perspective of Gyges, he may easily calculate the overwhelming likelihood that a perfect grasp would reveal an immeasurably greater value in the life. That life isn't Gyges' life, but in the perfect grasp of things that must define what Gyges really wants, all lives are equally included.

Let me briefly draw together some of the parts of what I've been saying. The nearly universal approach of philosophers to the question of the nature and authority of morality is to assume a contrast between, on the one hand, the concerns in individual systems of desire (including, perhaps, not just self-interested desires but also disinterested desires like those expressing disinterested benevolence and disinterested malice) and, on the other hand, the concerns of morality. Self-interest, they think, may be brought into line with moral concerns through threats of punishment and prospects of reward; but in circumstances in which these cannot operate it must seem impossible to give a reason to be moral, unless, perhaps, a motive of disinterested benevolence happens to be present.

This view contrasting individual systems of desire and morality represents the individual systems of desire as containing tough indissoluble cores. What I have argued is that, since desires must be thoroughly based on beliefs, systems of desire, down to the centre of their cores, are correctable with those beliefs. And the corrected, real desires belonging to all those systems simply *are* the concerns of morality.

I'd like this section to serve as an introduction to the fuller discussion that follows. So let's start again.

The Principle of Best Action

People care a lot about being moral. Sometimes to be moral they will seem to be acting against their own most powerful desires. From what does morality get this authority over them? Answers to this question that point only to contingent motivations like benevolence, fear or habit, can obscure morality's claim and undermine its hold on us. But aren't all motives contingent, liable to vary with agents and their circumstances?

No. There is a necessary overriding motive, one required in the very logic of agency itself, a motive whose expression is therefore a categorical imperative that could be issued to all agents at all times. This universal *principle of best action* would say to each and every agent, "As far as possible, do what you would want if you had a perfect grasp of what you were doing."

The hypothetical *perfect grasp* mentioned here would be a perfectly accurate and complete perception and appreciation, possessed with perfect firmness and responsiveness. What you would want if you had such a grasp of reality (including a perfect grasp of what it is like for you and others not to have that grasp) is also what you *really desire* as you actually are.

If someone eats something not knowing it is poisoned, he is doing what he desires but probably not what he really desires, not what he would desire with a perfect grasp of what he was doing. And if someone yields to temptation and eats something he knows will make him sick, not because he thinks the pleasure is worth the sickness but merely from weakness of will, the desire on which he acts, if indeed the pleasure is not worth the sickness, fails to be a real desire. For the perfect grasp of what he was doing, which would have inspired his real desires regarding the action, would have had to be possessed by him with perfect firmness and responsiveness as well as with perfect accuracy and completeness. The sickness would have confronted the pleasure directly in that grasp and overcome the temptation.

But even an agent who thus through the weakness of his will knowingly acts against his real desires really wants to be acting in accordance with them. For there can be no self-interest in pursuing aims that the pursuer himself would reject if he had a perfect grasp of what they involved. And there cannot fail to be self-interest in pursuing aims that the pursuer would embrace if he had that perfect grasp. The reasons for action that can carry real

authority for him are all and only those reasons that would be found sound and compelling in a perfect grasp of reality.

Every desire is necessarily dependent on the desirer's grasp of its object and is therefore necessarily correctable with an improvement in that grasp. As in our earlier example, I desire to drink what I think is hot chocolate, but that desire changes when I realize it is mud. I could not credit any desire that I thought was based on a grasp of its object that was relevantly wrong (because the drink was mud) or relevantly incomplete (because the drink was chocolate, but it had arsenic in it). I could not credit, therefore, any desire that I thought was not consistent with a perfect grasp of all the possible objects of desire (or undesire).

Desires and actions that are not based on the real natures of their objects are simply not the real responses of the desirers and the agents to those objects. The point of the perfect grasp is that it is the ideal medium of responsiveness to reality itself. Only through that medium could reality itself become the basis of desire and action. And the self-interest of any agent requires reality itself to be the ultimate basis of all his actions.

There is a seeming exception to these claims that I think helps to prove the rule. There could be a real desire to have another desire that was not itself real. For example, a child's desire for an actual visit to the moon (as opposed to a fantasy visit) might be based on an ignorance of the visit's difficulties and nastiness but might also, as an unrealizable desire, be harmless and fun to have. Thus having the desire might be really desirable. But being an object of a real desire would not turn the moon desire itself into a real desire. For the object of the real desire would be the moon desire and not the moon desire's own object, the moon visit.

The point of real desires is not at all that they would have been desirable to feel, as the moon desire would be. Real desires would anyway only be felt in a hypothetical perfect grasp of reality. The point of real desires is rather that they, and only they, as based on that perfect grasp of reality, have power to define for us which objects are really desirable (including perhaps a mistaken felt desire to visit the moon).

From nothing more than such necessary truths about agency and desire springs the overriding motive of each and every agent to conform to the principle of best action, to do, as far as possible, what he would want if he had a perfect grasp of what he was doing. And, as we shall soon see, from this purely rational principle, in itself neither prudential nor moral but only expressive of one's real desires, flow both prudence and morality.

A Conversation with Rob

That morality was part of what one really wanted occurred to me for the first time during a conversation with a friend in the mid-nineteen sixties. Rob, as I shall call him, told me, without any shame, about how he was stealing comic books from the shop of his unlikable employer. There arose a challenge. Could anything be said against this stealing that could move a free spirit like Rob? As a start, I asked if he could give me the general principle behind his behaviour. He said something like this: I wouldn't want to be stealing from the people I like and care about, from my friends and my family; but outside of that I don't mind.

It struck me suddenly that this principle had something demonstrably irrational about it. I sensed it could have Rob doing things that in reality he didn't want to be doing. To bring this out, I asked Rob to help me analyze why he would not steal from our shared friend Bert. In his principle, of course, Rob had made his liking Bert the basis of his not wanting to steal from him. I now pointed out that we could distinguish the fact that Bert was a likable person from the fact that Rob had actually come to like him. And when we looked a bit more deeply

at his motivation, Rob admitted that it was Bert's likableness rather than his actually being liked by Rob that gave Rob his real basis for not wanting to steal from him. Rob said he was unhappy at the thought of anyone stealing from Bert quite apart from any stipulation as to whether Rob happened to be appreciating Bert at the time.

"So," said I, "if you had happened not to get to know and appreciate him, you could according to your principle be stealing from him. But then you would yourself be doing something that in reality you would not want anyone to be doing. Your principle allows you to be an enemy to what you really want." And Rob agreed. In his treatment of strangers he must try to be taking account of what in reality he wanted, not merely how he happened to be feeling about them.

But, as we soon noticed, an extension of Rob's concern from just the people he actually liked to the strangers he might also have found likable was still far too superficial to have brought him all the way to his real desires. The argument developed irresistibly further till it covered as well the unlikable boss. For the definition of Rob's best action could not have stopped with what he would have wanted with a merely somewhat better grasp of what he was doing. In fact, a little more knowledge, or even a lot, can be a dangerous thing. It could be better for Rob that he remain ignorant of a bag of money buried in his garden if what he would have done with that money would have ruined his life. The only actions that are immune to such danger, indeed the only actions that do not succumb to such danger, are those that would have been endorsed with a perfect grasp of reality. And surely there must be enclosed within a perfect grasp of reality much more than just one person's narrowly-based actual and potential likings and dislikings of others.

A perfect grasp whose object is a thing that has no consciousness, a rock for example, will not include a grasp of what it feels like to be such a thing, since that thing will have no feelings. But a perfect grasp whose object is a conscious being must include a perfect grasp of what it feels like, of what it *is* like, to be that being. And a perfect grasp, a perfect perception and appreciation, of what it is like to undergo an experience had by that conscious being can only be had as a perfect sharing or recreation of the experience.

The perfect grasp of reality must be as if everywhere, within all experience, if it is to serve the purpose for which it was invoked, as the ideal medium of responsiveness to all the real objects of desire and action. And this is how the perfect grasp acquires values; it must capture within it all the compelling character of experience to be grasping it fully and accurately. It thus must be empathetic.

But a perfect grasp of a plurality of experiences, while it must be empathetic, must also be critically evaluative of each experience in the light of every other actual and potential experience and all the rest of reality and its potential.

Let me take a moment to point out that in a world in which there did not actually exist a perfect grasp of that world, the perfect grasp would be merely hypothetical. And a hypothetical perfect grasp would have to include a recognition of its own actual non-existence. It would not be itself comprehended as a further element in the world it was comprehending. It would, with perfect perception and responsiveness, enclose within it everything in that world without imposing on it any fictional contribution of its own.

Anyway, as I shall explain more fully later, what must be desired in a perfect and therein empathetic grasp of all conscious beings would be the harmony of all their real interests.

Since Rob's real desires are those he would have had with a perfect grasp of reality, what Rob really desired was what he would have desired if he could have grasped the world at once from the position of every desirer, including the boss from whom he was stealing, feeling exactly what each felt, but also evaluating each perspective in the light of all the others and of all they had missed of the world and themselves and their potential. Such an

apotheosis of desires, into the desires of an omnipresent, omniscient being, represents no more than what any desirer of any sort must really be wanting.

We started with Rob's responding to Bert, his liking him, and we tried to make this apply with more consistency to the full reality affected by his actions. Now I am saying that Rob is forced by the same pressure in the same direction to a point where he respects equally the real interests of all desirers, as these interests would be revealed in a perfect grasp of all reality.

But what if Rob had been so shallow that he was unmoved by the first point, about the possible likableness of strangers, and had insisted that he really only wanted to guide his behaviour according to the pattern of likes and dislikes that he happened to possess at the moment of his acting? Let's deal with something even harder. What if Rob had been a rock? Real desires are those that would be had with a perfect grasp of reality. There are therefore real desires belonging to a rock, namely those it would have had with a perfect grasp of reality. The real desires of a rock, or of the shallow Rob I just described, or of Rob as he actually was, or of any saint or sinner, or of God, the real desires of each and every person or thing must be conceived of as exactly the same, because the perfect grasp of reality must be the same, in its perfection, no matter from which angle we think of it being approached. The perfect grasp contains all the angles.

So what about stealing those comic books? If Rob was to choose according to the reality of what he was doing, one part of the reality to which he had to be responsive was the part that lay behind the boss's eyes. If the victim in a cold-blooded mugging that has netted a few pounds for the mugger has been killed, there is a special horror in the disproportion between the action registered as a small gain behind the killer's eyes and the vast loss that it is behind the victim's. Stealing comics is not so momentous and what it is for the victim as a loss is more in balance with what it is in the agent as a gain, but in a perfect grasp of it any significance of the action for the victim, in his thoughts or feelings, must be fully comprehended. And, of course, the perfect grasp must include much more than what the boss himself might experience. It must include a perfect comprehension of the reasons he is the way he is and of his potential for changing if he is treated differently.

And the perfect grasp would not, although it must include an empathetic hold on the boss's limited consciousness, constitute an endorsement of the boss's non-real desires any more than it was an endorsement of Rob's. What it would endorse would be the real desires of the boss, what he would want were *he* deciding this with a perfect grasp of the reality of others and himself and of the way the world worked. The principle of best action requires the sort of reconciliation of all wills that would be formed in a perfect, and therefore merged, grasp *by* everyone *of* everyone and everything. Doing what is right for oneself turns out to be doing what is right for all those affected by one's actions. Self-interest, transmuted through a principle that simply makes it consistent within itself, turns out also to be morality. This, though not in so many words, is what I ended up saying to Rob; and at least as I remember it he agreed. Stealing comic books from his boss was most likely not what he really wanted to be doing. Therefore it was wrong, it was irrational, it was a mistake. He had a reason, an ideal, overriding reason, to act with regard for the interests of all, whatever he actually felt about his boss.

The Principle in Action

The tension, between what an agent may feel he wants to do and what that same agent really does want to be doing, what he would feel he wanted himself to be doing if he were grasping perfectly the nature of what he was doing, this tension gives rise to the pull of morality and prudence in the motives of an agent capable of sensing the distinction between

its current consciousness and the whole of reality. (Morality and prudence have been revealed to be two characterizations of a single set of real desires. Morality is in their horizontal dimension, their responsiveness to others; and prudence is in their vertical dimension, their responsiveness to the future. Prudence is usually taken to be a self-interested farsightedness that will often conflict with morality. But morality and real prudence can never disagree, because morality, like real prudence, must agree with real self-interest.)

It must always be best for the agent himself that he do that which, if he were grasping perfectly what he was doing, he would want himself to be doing, that he do that which agrees with his own real desires. Whatever his more immediate motives may be, to the extent that an agent is rational about his actions he will make his actions conform to his vision of this ideal, of what his actions *should* be.

Notice, by the way, that it is not rationality as such that is here being recommended. It might be a good thing in some respects not to be rational from time to time. But if we are irrational in assessing what is best, if our motivation is incoherent, that can prevent us from doing what is actually best for ourselves, what we would want ourselves to be doing grasping perfectly what it was we were doing. Rationality determines coherence, and what is important here is coherent motivation. That is the primary value of rationality for agents, though it has many secondary uses too. But rationality is not being worshipped here.

Now let's ask, should I scratch my badly itching sunburned back with this backscratcher I am holding in my hand? A perfect grasp of possible experiences such as those that are relevant to this decision about scratching must include a comprehension of what those experiences would be like for the possessor (or possessors) of them. This part of having a perfect grasp of such possible experiences must be subjectively just like having the experiences themselves.

The perfect grasp of everything relevant to this decision about scratching must include at once a perfect grasp of the way it would be both if I don't scratch and if I do. It must include the full frustration that goes with not scratching and also the great but brief relief from the itch that I'll have if I do scratch as well as the terrible pain that would follow that relief while the itch was returning.

Such a perfect grasp would thus have to comprehend at once, and perfectly, states of consciousness that essentially exclude one another. Perhaps this means that our hypothetical perfect grasp of reality is logically impossible. But, possible or not, omniscience is the inevitable ideal of our knowledge and the perfect grasp of reality is the inevitable hypothetical basis of an appropriate responsiveness to reality, which is the whole point of action. The perfect grasp need not be logically consistent to have this significance.

Let me mention a further feature of the perfect grasp that may make it impossible. Even as the grasp is perfectly comprehending the possible future experiences of scratching and of not scratching, it must comprehend also that these are as yet merely potential experiences and that only one could become actual. And the perfect grasp must appreciate fully that what is important in the actual world is that the better possible experience become actual. In other words, although each potential experience must be fully grasped with all its compelling character, it must still be valued with regard to whether it should be made actual through action, as though the experience has not yet been had, as, indeed, it has not.

In my actual limited state of consciousness, of course, I shall experience the pain only if I first have had the relief of scratching, or else I shall experience the frustration of not scratching but also be spared the pain that would have followed scratching. And let us just say that in the perfect grasp of all these experiences I would have found that the relief from the itch was not in fact worth the pain that would follow it. This hypothetical judgment

within a perfect grasp would, then, define my best real action, or in this case inaction; I should refrain from scratching.

Notice, by the way, that it would be ridiculous to regard the possession of the perfect grasp as a particularly desirable state to be in. In the actual, limited state of consciousness I happily have the option of avoiding any close acquaintance with either the frustration of not scratching or else the pain that follows scratching. The perfect grasp, however, must comprehend fully all the things I might want to avoid, combined, perhaps impossibly, with a full appreciation of all the escapes from them that might be had.

So the principle of best action is badly misunderstood if it is taken to be recommending an *actual* perfect grasp or an actual feeling of one's real desires, which would anyway be unattainable for us and perhaps contradictory. Neither does the principle, considered in itself, recommend truth, knowledge or an increased perception, appreciation or grasp of things in any degree whatsoever. Though an increased grasp of things will often be useful in deciding which actions are endorsed by the principle, and in carrying them out, and some increases in grasp would be discovered in a perfect grasp to be inherently desirable, the purely hypothetical perfect grasp that is mentioned in the principle is merely employed for its aptness in defining the best course of action for an actual, limited consciousness.

An ideal grasp of reality only lends itself to defining ideal actions because having a grasp of the sort here invoked would constitute being fully and correctly responsive to reality. It is this responsiveness to reality and nothing else about the perfect grasp that gives it such relevance to action. Let me therefore paraphrase what I earlier said about rationality: perception, appreciation and grasp are not worshipped here.

In fact, though a perfect grasp is necessary in the hypothetical state that would define best action, knowledge that is merely propositional will often be far preferable as a basis for the actual carrying out of such action, as I shall now explain. Returning to my itch, let's say that I with my ordinary state of consciousness do know propositionally that "the relief from the itching would not be worth the pain", but I find that I still want to scratch because the insistence of the itch is exerting a disproportionate force in my actual state of consciousness at this crucial time of my decision. The possible pain is now not fully grasped; it is now not fully perceived in all its power; and possessing the propositional knowledge that "the relief from the itching would not be worth the pain" may not be all that I need to make myself behave appropriately. A genuine perfect grasp of a pain, since it would contain within it the pain itself, would have all the motivational force of that pain. By contrast, the mere entertaining of a proposition, which only refers to or describes the pain with its words, will not *in itself* carry such motivational force.

Yet there is a way I could try to bring to bear on my actions at least some of the absent force in the perfect grasp. I could try to oppose the current insistence of the itch with a vivid imagining of the pain I would be feeling soon after scratching. I would be creating an image of a missing content of the perfect grasp of what was relevant to my choice so that this missing content might yet throw some of its weight against the actual temptation in my current perspective.

But rather than to attempt such unpleasant efforts of the imagination, it should be open to me as a rational agent simply to allow the mere proposition that "the relief from itching would not be worth the pain" to function in my motivation as though it were itself the perfect grasp of these experiences. For if such a proposition can claim the special authority of representing reality as the basis of my actions, it may enter among my motives like a referee, slight yet capable of settling a fight among heavyweights. And perhaps, because I am rational, I can behave as though motivated by a pain that I do not have.

I claim that the same kinds of consideration apply to a decision about whether I shall hand the backscratcher to a child suffering from a similar itch. But in this case I should act

as though motivated by a pain I not only do not but will never have because it is another's. In a perfect grasp of the arena of my actions that pain would be assessed in the same way as one of my own. Although any boundaries of persons would be perfectly grasped, that what I had called "I" lay within only one set of these boundaries must be unimportant in the perfect grasp. "I" just refers to one among the conscious beings each of whose full reality would be equally comprehended there. Who I am in my ordinary consciousness is, as would be discovered in a perfect grasp, not relevant to my real desires.

Let me say more about this crucial and perhaps difficult point. I am not here saying, though I think it is true, that if I were not limited in the reach of my consciousness I would also not be limited in my identity. What I argue for elsewhere¹ is even stronger. I argue that there is but one possessor of all consciousness, of whatever reach. This one subject of consciousness mistakes the merely epistemic boundaries of limited reaches of consciousness for metaphysical boundaries of distinct conscious beings.

The same mistake is clearly exposed in the case of brain bisection. Consider that if you became a split-brain patient, someone in whom the connection between the brain hemispheres had been surgically cut, each of your then unconnected fields of consciousness would falsely appear to you to be all of your consciousness. Within the reach of consciousness in each hemisphere you would appear to yourself to be a different person from whoever was conscious in the other. I argue that I, the universal subject of consciousness, am similarly deluded into thinking that the limits of each among all fields of consciousness are the limits of me. But what really makes consciousness mine, and the person having it I, is simply the first-person quality of experience, its immediacy for whatever is having it. All consciousness has this quality. Hence all consciousness is mine and all persons are I.

If what I mistakenly take to be uniquely my consciousness could be expanded to include the content of all the others, as it would be if I were given the perfect grasp of reality we have been discussing, I would be discovering vividly that all the content I had thought belonged to others really had that same first-person immediacy that had marked as mine the content of the limited consciousness with which I started. The expansion would not be making it mine. The expansion would merely be revealing to me that it was mine, just as a reconnection of my hemispheres after brain bisection would be merely revealing rather than bringing it about that the consciousness of both hemispheres was mine.

I think that, at the deepest level of understanding, my thesis of personal identity and my thesis of morality are joined. A perfect grasp of reality must reveal the compelling character of all consciousness for me because it reveals that all conscious states are mine. But I don't need to insist on this. And I shall now return to my point that real desires would not vary despite a distinctness of persons. Although I will happily concede that, on the assumption that persons are distinct, a person with a hypothetical perfect grasp might be expected to know which among the actual ordinary consciousnesses was his, I maintain that this knowledge could never motivate him with his perfect grasp, since that grasp necessarily contains all of what is in all the consciousnesses affected by his actions.

Notice that a person in his hypothetical perfect grasp of reality could not for the purpose of defining ideal action be thought of as somehow advising himself in his actual limited consciousness on how he could bring about the greatest possible gratification within those narrow boundaries. For it is only *as* perfect, with the very motivations that the perfection gives it, that the perfect grasp can represent what a person, even with his limited consciousness, *really* wants to be doing. And in a perfect grasp of more than one

¹ The "One Self: The Logic of Experience", *Inquiry*, vol. XXXIII (1990), pp. 39-68.

consciousness the wants cannot centre on satisfactions within the boundaries of just one, no matter whose it may be.

But can we really be sure that a person's recognition, with a perfect grasp, of which limited consciousness is his would not undermine the impartiality I am claiming for the possessor of a perfect grasp of reality? If, as I've mentioned before and am about to explain further, there must be a different mode of evaluation of potential as opposed to actual experiences, perhaps a perfect grasp must also include a dramatic difference in evaluation when comprehending its possessor's limited consciousness and that belonging to another.

Actuality must always be grasped as the focus of all concern in the perfect grasp of reality. Although the person with the perfect grasp must be perfectly grasping every merely potential experience, he must somehow not be responding to it as though it were already actual; he must be responding to it as what it really is, one alternative among the ways that actual experience might and might not develop. (Perhaps this requirement, as I mentioned earlier, represents another arguable inconsistency of omniscience.) The person with the perfect grasp must grasp perfectly what it would be like to be having a potential pain and yet respond to it with the desire to avoid its becoming actual, as though he could still keep a distance from it.

But, then, why could we not have the person with the perfect grasp also fully grasping what a pain is like for another but responding to that pain as though he was at a distance from it, as though it had for him, even as perfectly grasped, a different kind of value from his own pain and so could be ignored or even cruelly enjoyed?

Before I argue against such discounting of others in the perfect grasp defining real desires, let me point out that a full grasp of everything beyond one's own limited state of consciousness, even if elements in this could be discounted, would yet be like an earthquake to the unenlightened motivation resting on distorted appearances in the ordinary consciousness. It is plausible to think that not much cruelty would survive a grasp of reality that included a full empathetic understanding of a possible victim, even if, because of discounting, the grasping of the other's full reality was not to be thought of as automatically compelling, as containing within it for the grasper the very force that it has for that other.

Shelley Kagan argues on a basis like this for the rationality of acting with an "extreme" regard for morality. He appeals to "the claim [that] I would tend to act in keeping with the objective standpoint were all my beliefs vivid" and "were I [therein] a better judge".² As Kagan himself points out, there is a tradition of such arguments running from Plato to Brandt. The philosophers of this tradition claim that morality can be defined in terms of the interests one would have had with some sort of ideal awareness. One problem for such a view is establishing that the interests inspired by an ideal cognitive state would indeed agree with morality. Kagan, since he imagines an ideal awareness that is like our ordinary awareness in that it allows what I have described as a discounting of the interests of others, argues merely that it is unlikely that someone informed by vivid beliefs about the condition of others would care to exercise his option of ignoring their interests.

A second problem for such a view is explaining why the actual self-interest of someone without an ideal awareness should be influenced by the interests that he might have acquired if he did have an ideal awareness. Attempted solutions of this problem have usually depended on those features of normal self-interest that could make us respectful of the verdicts of greater knowledge or sensitivity, such as rationality, prudence or benevolence. Kagan, as in the quotation above, appeals in part to a common acceptance of the proposition that an improved awareness would make an agent in some sense a "better judge" of what he

² Shelley Kagan, *The Limits of Morality* (Oxford: Oxford University Press, 1989), p.302.

should do. (And if being a better judge can mean something like judging better what one really wants, this seems to me at least a gesture in the direction of what I am saying.)

These two problems of the tradition are closely related. The more the ideal state of awareness is thought of as belonging to a detached ideal observer, the closer its interests are to morality. But that then makes it harder to tie these interests to what is conceived of by these philosophers as the distinct self-interest of the ordinary individual.

I have been arguing that the distinction drawn by these philosophers between one's actual self-interest and the interests one would have had with an ideal awareness is mistaken. Anyone's real self-interest is identical with the interests that he would have had with a perfect grasp of reality. Furthermore, as I am now about to argue, this perfect grasp defines morality as well as self-interest because it would not allow a discounting of the interests of others; the real self-interest of anyone is also the real self-interest of everyone else.

For although there is good cause to evaluate differently merely potential as opposed to actual experiences within a perfect grasp of reality, since that which is merely potential is not yet real, there can be no similar cause for any such special evaluation of others' experiences as opposed to one's own, since others' actual experiences are as real as one's own.

That others' experiences are not one's own would, as I earlier said, be fully recognized in a perfect grasp of reality. But only in an ordinary consciousness, due to the limits of its perception of reality, does such a distinction mark any failure to experience the compelling character of another's experience. Why should it have any such significance in a perfect grasp of reality that defines real desires? Isn't it the very sort of thing that makes the perfect grasp supreme in determining one's responses to reality that it would cross the epistemic boundaries between consciousnesses to show what is really inside them all, with all the otherwise unfelt power of their motives?

The distinction between self and other as recognized in a perfect grasp could never inspire any discounting of the other, fully contained there as the other is. Quite the contrary. The perfect grasp can only inspire a perfect attachment to the real interests of the other. A supposedly perfect grasp of reality that had discounted the experiences of others would therein have ceased to be a perfect grasp of reality and to define our real desires. To the extent that one applied such a discount to another's interests one would end up doing things one wouldn't want to be doing with a perfect grasp of reality, one would fail to be responsive to the full reality of what one was doing.

But imagine a doubter now speaks as follows: Look, we are trying here to develop an ideal of self-interest. Doesn't this give us the best possible reason for designing the perfect grasp that is to define ideal self-interest with the very feature you call arbitrary, the discounting of others' experiences and desires relative to one's own?

My answer is that the resonance of the term "self-interest" has confused this doubter. He is used to thinking of self-interest as something revolving around the prospects for happiness within the actual limited consciousness belonging to that one self whose self-interest this is supposed to be. But in speaking of self-interest we have not been talking about that kind of happiness for any one person; we have been talking about a person's real desires, about what would make him happiest if he were fully grasping all that was really going on. This may well involve a sacrifice of happiness he could feel himself in his ordinary consciousness. We have been talking about self-interest of the deepest sort, a sort that would reject a shallow happiness that stood in the way of responding fully to reality. And a perfect grasp that would define this real self-interest could not discount any part of reality.

Let me briefly discuss a related point. A perfect propositional knowledge of everything that could be relevant to your choices, as opposed to the perfect grasp we have

been describing, would be inadequate for defining your real desires and best action. As we have seen in the case of the itch, you can act against even the best of judgments when they are in the form of propositional knowledge; this form does not in itself have the motivational force of the reality it represents. Only the sort of perfect grasp I have required could be certain to render you perfectly responsive to the reality of your actions.

What I really want to be doing in the case of the itching child, then, is what I would be forced to recognize as desirable if I were stuck within a perfect grasp of all of the actual and possible experience in the child's consciousness and my own and whatever else could be affected by my decisions. At the heart of this perfect grasping would be an adjustment of the values contained in the limited consciousnesses to one another, a sorting out of which reactions in them were appropriate to the rest in the light of the full grasp of them all. This would not be like a passive collection of these states; it would be a creative discovery of the fittest plan for them.

But what if I know that in my limited state of consciousness I'll get sadistic pleasure if I hand the backscratcher to the child and then see, from the distance of my own perspective, the consequent pain? How should I regard this potential pleasure? What should I do?

Well, in my perfect grasp of reality I would have to be perfectly grasping, with all its motivational power, the sadistic pleasure that could exist in my limited consciousness. But I must also be perfectly grasping, with all its motivational power, my victim's pain, despite that pain's belonging to another. And, since the sadistic pleasure would have had to depend on the pain's not being felt by me because it belonged to another, we might easily think that this double grasp, of sadist's pleasure and victim's pain, would simply annihilate the sadistic pleasure. It might seem, then, that a desire for sadistic pleasure must be inconsistent with a perfect grasp of reality, that a desire for such pleasure, since it could only arise within a stunted consciousness of reality, one that shut out the victim's perspective, could never be a real desire.

But it's not that simple. What if I sought a little pleasure, perhaps as a distraction from my sunburn, in attempting to solve a crossword puzzle? Such pleasure must depend on initially not knowing the puzzle's solution. Since that solution would be fully grasped somewhere in a perfect grasp of reality, a person with the perfect grasp might seem incapable not only of valuing sadistic pleasure but also of valuing the rather innocent pleasure of filling in a crossword puzzle. Must the attempting of a crossword puzzle be condemned because it seems in this way inappropriate to a truly comprehensive state of consciousness? But the perfect grasp must be a perfect grasp of all the values within the limited grasps. Indeed we started our discussion of sadism with the point that a sadistic pleasure, with all its motivational power, would have to be perfectly grasped within a perfect grasp of reality; and we may now add that the pleasure of attempting a crossword puzzle would have to be perfectly grasped there too.

What we see here is a tension between two elements in the ideal of a perfect grasp of reality, a tension that arises from a seeming contradiction in omniscience that has figured importantly before in our discussion. On the one hand, a perfect appreciation of reality must include a perfect empathetic grasp of actual and potential states of ignorance and distorted appreciation; and so it must include the pleasures of both sadism and crossword puzzles. Yet, on the other hand, the perfect grasp must also include an over-arching final assessment of these states in the light of full reality, including the reality of victims' pains and the crossword puzzle solutions. And we may find ourselves emphasizing one or another of these aspects of a perfect grasp, either the identification with the alienated limited states of consciousness or the over-arching assessment of these.

Here it is important to recall a point I made earlier, that a merely hypothetical grasp of reality must include a full grasp of its own actual unreality. If the person's perfect grasp is merely hypothetical, the stress must be on the reconciliation of the only real states of consciousness, the limited ones, rather than on any hypothetical gratification of that person in his non-existent synthesis of these states. And this is where a religious morality may show differences with an atheistic, humanistic morality. If an omniscient God exists, then the over-arching final assessment is an impressive state of awareness that actually exists. In that final assessment sadism would indeed be impossible and crossword puzzles senseless. Real desires that had to take account of such an actual all-embracing mind might not be tolerant of either the meanness or the frivolousness of merely human concerns. But real desires as envisaged by an atheist, or by a theist who emphasized God's own engagement with and tolerance of limited consciousnesses, would be more liberal about such matters.

Yet there would be reason enough for even the liberal perfect grasp of reality to favour crossword puzzles more often than sadism. Sadism is like a taste for sweets; its pleasures may have a cost in pain. The greater such pleasures are, the more dangerous they may be. A perfect grasp of reality would fully endorse only those sources of pleasure that best promoted the reconciliation of all the real interests contained within it. Only a sadism that could be safely restrained or channelled so that it would agree with all these real interests could be found desirable within a perfect grasp of reality.

But maybe sadism should get a much stronger endorsement than this from a perfect grasp of reality. After all, wouldn't the course of action most desired by someone in the largely unpleasant condition of possessing a perfect grasp be to escape it and to adopt instead the most favourably positioned limited consciousness available? And in that best limited state of consciousness surely he would be enjoying the pains of others rather than having to share them, as in the perfect grasp, or even caring about them, as a moral person would. My reply is that this once again misses the whole point of invoking the perfect grasp of reality. (I say "once again" because this is a complaint we heard before, though in a somewhat different form, when we considered whether you with a perfect grasp would not just want what was best in the narrow consciousness you knew to be yours.) Imagine that you with a perfect grasp could shrink back to possessing only a limited consciousness, perhaps taking with you just the information you would need to make your life a satisfied one within that narrow compass. Well, after the shrinking your best action would *still* have to be defined by what you would have been deciding with a hypothetical perfect grasp of everything affected by your choices. The knowledge brought back to your limited consciousness could be a dangerous thing without reference to the perfect grasp of what it really was you were doing with that knowledge. Your real desires, the things you really want to be doing, are what that perfect grasp, and only it, represents. The point is neither to have that grasp nor to escape from it, but rather, as far as possible, to conform to what it would have had you wanting.

Let's look now at the various ways we might fail to conform to the principle of best action. I might want to do the right thing for the child and yet do the wrong thing by giving it the backscratcher because I do not know that the scratching will result in pain or I do not know that the pain will be worse than the relief from the itch. I might be either mistaken or consciously ignorant about this, with more or less justification. While we dislike the consequences in such a case, we may still respect the agent's good intentions. Yet much of the horror of history has followed from such mistakes or ignorance combined with good intentions.

There have been two mistaken beliefs about the nature of morality itself that have done their share of mischief. Morality has too often been accepted as a bundle of rules to be stiffly applied rather than an attempt to be responsive to reality. And, sometimes partly in

reaction to such arbitrariness, morality has also been cynically rejected, as not having any legitimate claim in a person's rational deliberations, although it might at points be acknowledged to happen to agree with his sentiments or his narrowly conceived self-interest, immediate or long-term.

Another mistake about morality could be a fanaticism in the application of the principle of best action itself, having us always concerned about distant others and a distant future. If no special weight were given to oneself and those close to one, or to the present moment, much of human life, including that of similarly concerned others and future moments, would be poisoned, or starved. Thus a moral concern for others may best be tempered by a particular concern with one's own affairs. A prudential concern for the future may best be tempered by a respect for spontaneity and living in the present. Anyway, we often know more about ourselves and the present and can deal better with them. But none of what I been saying should be seen as a qualification of the principle of best action; it is what the principle itself must recommend. The principle has the power to absorb like a sponge all criticisms according to which it would be recommending anything undesirable, such as the self-defeating fanatical concern for others and the future that we have just rejected. The hypothetical perfect grasp itself must contain an appreciation of all such problems and any solutions. And it would include, as part of this, a full recognition of the limits and special needs of particular agents.

So far I have been describing ways one might fail to conform to the principle of best action because of ignorance or mistaken belief. But there also can be weakness of the will. As I described earlier in the case of my own itch and still earlier in the case of an agent eating food that he knew would make him sick, one might yield to a present temptation and know this is wrong. It is possible to possess the propositional knowledge that the pain that follows scratching will outweigh the initial relief and yet hand the backscratcher to the whimpering child; and one may do this even though one desires to do what is right for the child. For I might act to help the child to escape its immediate suffering though I know that this is at the cost to the child of more suffering all told than it would otherwise have had to bear. What is most missed here, of course, is the perfect firmness and responsiveness with which a perfect grasp must be possessed rather than its accuracy or completeness.

If I have knowingly yielded like this to the disproportionate power of a current temptation, my conscience will bother me; the opposing and proper pull of reality on my motivation, the claim of that which I really want myself to be doing, will in my conscience become palpable to me. And in an attempt to relax this unpleasant tension I might yield to another temptation and indulge in self-deception, in rationalization, persuading myself uneasily that I *am* doing the best thing I can.

There may seem to be yet a third kind of failure to conform to the principle of best action, that of the agent who is not in any way ignorant or mistaken about what is right but who turns without conscience against it, to pursue a narrowly conceived self-interest or even just through wilfulness. But does this category actually make sense? Agents who were evil without conscience because they were ignorant or mistaken about the claims of morality itself, who perhaps thought that morality was really a trick played by society on the gullible or that morality was all right for the naturally benevolent but could be rejected by them because it was simply not to their taste, those who were evil without conscience because of some such general ignorance of or mistake about the character of morality itself, surely still belong in our first category, of those who fail to conform to the principle of best action because of ignorance or mistaken belief. And anyone, it seems to me, who really knew the nature of morality, who really accepted that morality represented his own real desires, would be unable to adopt the sort of swaggering attitude towards it that would be required, along with such knowledge, in the putative new category.

Non-consequentialism

The principle of best action requires an agent to be responsive to the truth about the world and his own nature, but the principle itself says and assumes nothing concrete about that world and agent; the principle is derived merely from the abstract idea of agency, which is why it is categorical, unconditional. It is the principle plus an idea of the world as containing various sentient beings and times that gives us morality and prudence. In a dream or in the world as imagined by a solipsist, who sees himself as alone with his experience, the principle would still apply but would generate no morality, since morality is concerned with others.

The morality I have so far derived has been consequentialist. The principle has simply made us attend to the interests of everyone affected by our actions as these would be adjusted to each other in a perfect grasp. But what if, as Kant believed, a perfect grasp of reality would reveal that the metaphysical character of the rational agent obliges him, with categorical force for such an agent, to act with the autonomy and dignity that comes with regarding only the form of his actions rather than the consequences of them? The principle of best action when applied to the Kantian view of the nature of the agent, then, could yield a formal or deontological rather than a consequentialist ethics.

But Kant didn't invent deontological values. They pervade our ordinary thinking. An example of enormous importance because of the misery it generates is our retributivist attitude towards punishment. It is based, I shall argue, on a natural mistake in our thinking about ourselves.

Laplace famously claimed that, if he knew precisely the present state of the world and the natural laws, he could in principle predict all the future development of the world. But, even if we were to grant Laplace's stipulations and assumptions about the world and its laws, the prediction itself would be in principle impossible. For in order to determine the effects on the world of his own prediction he would have to determine first the precise character of that prediction. He would therefore have to have determined the prediction before he could make the prediction. The absurdity of this leaves us with an inevitable indeterminacy of prediction which is based in the indeterminacy of *self*-prediction.

It would be possible for one computer to predict with precision a future state of another computer by calculating that future state from precise information about the other computer's present state and the rules that will govern its development (plus information about future external influences). But that first computer could never do a similar sort of job of precise *self*-prediction. For it could never have calculated its future state without first having calculated the intervening state in which it would be making its prediction of that future state. But that would mean that it would have to have calculated its prediction before it could have arrived at its prediction, which is an impossible requirement.

Popper and others have pointed out that it is impossible for a computer to represent its own current state with precision and completeness. For within the representation it would have to be representing the representation itself and also the representation of the representation and the representation of that and so on to infinity. Yet there would be no such problem in a computer's entertaining a precise and complete representation of its own past states or another computer's present state, since the representation of such states would not be required to be representing itself. And the computer could be describing its own present state or predicting its own future states, but always without precision or certainty regarding the part of itself engaged in self-description or self-prediction. (Here, by the way, would seem to be another unworrying problem for the unneeded consistency of the perfect grasp of reality, except that, as I earlier explained, anyone who possessed a merely

hypothetical perfect grasp would be perfectly unconcerned to know himself *as* possessing a perfect grasp since that is no real part of the world that this hypothetical state is designed to be comprehending.)

Now we come to our natural mistake in thinking about ourselves. Each of us strongly feels the impossibility of ever knowing fully or with certainty what he is or what he will do. If I try to grasp what I am or will be, something of that which tries to grasp, the grasping itself, must remain outside the grasp. Earlier I briefly described what I think is our illusion of taking mere epistemic boundaries for metaphysical boundaries of personal identity. In the case I am now discussing we are inclined to take the epistemic indeterminacy of self-prediction and self-description for a metaphysical indeterminacy of what we are.

In just such a way does Sartre elevate his apt metaphor, in *Being and Nothingness*, for our experience of indeterminacy to the status of a contradictory metaphysical truth. He accounts for the indeterminacy in self-knowledge of the present and future, which, as we have seen, we share with every computer, by insisting dramatically that as subjects of consciousness we are radically special, free of the fixed natures of mere objects and of ourselves in the past (in which our natures *are* fixed). Yet for Sartre this somehow does not properly mean that these subjects lack a nature. For they, after all, make choices according to their projects. These subjects must therefore possess in Sartre's metaphysics a contradictory nature without a nature.

I think we should view an agent as simply having a nature from which he acts. We could add a dash or more of metaphysical spontaneity, but this would not be helpful either in expressing our peculiar experience of ourselves or in providing us with a deep enough responsibility to support a retributive attitude towards us. The impossibility for an information-gatherer that it come to know itself completely or predict its own future with certainty misleads us philosophers into flirting with a metaphysical freedom from causes as its explanation. But that certainly doesn't fit with the way we make choices, usually in patterns and with purposes. And it would ironically dispense with *any* sort of responsibility for the uncaused choices and actions. It takes the subtlety of a Sartre to do some justice to our experience in his still unhappily metaphysical depiction of it, but we need the truth about a matter so important.

Its importance for us lies in the way we judge ourselves and how we should be treated. Think of an evil agent, one who deserves punishment. Think of the man who put pieces of glass in baby food in order to extort money from the baby food company. (What I shall next ask you to imagine, this man's having been manufactured by a mad scientist, could strike many readers as arbitrarily committing us to too scientific a view of a person. Please be patient. I'll soon be broadening the discussion to take in other views of how a person might be created.)

Now imagine you become convinced that this evil man had been assembled, the night before he launched his disgusting scheme, by an advanced sort of Frankenstein, a mad scientist capable of constructing from the raw materials of living things not a mere automaton, but a completely human-like being, with a brain supporting a mind that has in it the same pattern of conscious and unconscious mental life that we would have assumed this man had when we started thinking about him as our example of an evil agent. Since we have stipulated the internal sameness of this man with the man as he would be in the normal case, he himself will not know he was manufactured by the scientist. He has a complete set of memory impressions of having been a child, of a bank balance, etc., as well as all the same motivation he would have had in the natural case. So he does the awful deed.

Now what do we think of him? My reaction is to think he is bad and ought to be stopped and, if this is possible, changed. What I no longer seem to feel (as I confess I did before) is that he deserves to suffer, great pangs of guilt as well as what we decent people

will do to him. I no longer feel that there's an intrinsic good, a good apart from such possible consequentialist goods as reform or deterrence, in his feeling the pain of punishment because in some deep sense he *deserves* the pain. I see now too clearly for that that he is merely a product, with a nature chosen not by him but his manufacturer. I might just add, recalling what I have argued in this paper, that this agent was setting out to act against his own real desires in his stupid scheme. It is good *for him* that he was caught before he could have done more of what he wouldn't want himself to be doing if he perfectly grasped what it was he was doing. Anyway, when I consider my reaction to thinking of this agent as a product of the mad scientist, it seems to me that my usual retributive feeling has drained out of me.

Someone might be tempted to feel that the scientist deserves the punishment for choosing this evil nature for his creature. And, just for the fun of it, I could next ask the one who thought of this to imagine that the scientist too had been, without his knowledge, similarly created the night before he did his work by yet another such scientist. But this just delays our coming to the real point: what possible moral relevance can there be in the difference between the normal, actual case and this science fiction one? In contrast to our artificial man, the natural man in the actual case was produced by a less organized set of causes--heredity, environment, maybe initially God. But he, just like our artificial man, simply must act according to what he is--according to a nature which is internally indistinguishable between the two cases.

If we try thinking that, be he natural *or* artificial, some of his actions may have been somehow uncaused, this would only threaten to remove from him even the sort of responsibility that bad weather may have for a bad harvest. Yet we are all vulnerable to the illusion that a thinking being cannot really be limited by his own nature--that he is always free to rise above it and improve upon it. Of course, thank God, a person *can* resolve to improve himself and can *do* so, but surely this will only happen when *that* was in his nature, in that very self-reflection that must seem to the agent to be indeterminate.

I contend that no agent can be conceived of that could ever be responsible in the way required by retributivism. First, no agent could be responsible in any sense for his own beginning. In particular, he could not before he existed have chosen his own nature. Whether he was formed by God or things of the world or arose spontaneously, somehow free of any cause, *he* could not be responsible for his own original nature. And even if, like God in the ontological argument, he existed with necessity from his own nature alone, he would have had no choice in either his existence or his nature.

Next, only as far as he himself shapes his later nature can he be thought responsible in any sense for that, but such self-improvement or self-corruption cannot make him responsible for his nature in the way required for deontological desert. For, apart from either external causes or a metaphysical spontaneity for which nothing could be responsible, this self-determining can only depend on his original unchosen nature or later developments of that.

Finally, the agent cannot be held responsible in the required sense for any of his actions, whether these are thought of as determined by the nature he essentially did not choose or as bubbling forth somehow undetermined and thus with nothing responsible for them in any sense. I conclude we should cure ourselves of this monstrous mode of judgment that has caused such great suffering in the bitterness of the judge as well as in the pain of the needlessly punished.

But what of the Kantian view, already mentioned, that the agent when he acts morally, since he is then purely rational, can be free of all the particular, contingent conditions that seem to have formed him? Kant certainly thought that this sort of autonomy invested an agent with the responsibility required for deontological desert (although it seems

to me this could authorize only reward, not punishment, since to be rational and autonomous was to be good). I think this fails, because if the agent's motivation is somehow fixed timelessly by the forms of rationality, this is not freedom but necessity--it involves no choice. Neither could the agent, in the formation of his nature, have chosen either whether he was to be capable of rationality or whether he was to be disposed much to use it. (These same points apply to the idea that what we are entitled to punish retributively is the evil nature of the agent. Anyway, we can't either punish or sensibly blame for being evil the *property* of being evil.)

Of course, Kant at this point (inspired partly, I believe, by the impossibilities of full self-knowledge I have described) might appeal to his difficult distinction between the phenomenal and the noumenal. As noumenal the agent is beyond the limits of both causation and his own comprehension, though somehow the agent can feel a reverence for the moral law that is expressive of his noumenal freedom. I shall leave this discussion here, except to remark that it seems to me that an agent only possesses autonomy and dignity in a straightforward sense when he respects the authority of reality as the basis of his actions in conformity with our principle of best action; and this autonomy and dignity does not derive from a problematic metaphysics but from the analytic character of agency.

Deontological ethics might have another source. It might result from a strong emphasis being placed on the supposed boundaries of persons. A libertarian philosopher like Robert Nozick will insist that it must be wrong to think it an enforceable moral requirement that an agent concern himself in a positive way with the well-being of others. Proper attention to personal boundaries, he would argue, must reduce enforceable moral requirements to obligations of non-interference in the lives of others.³ Liberty that allows others their liberty is the absolute right of each.

The position of John Rawls can be seen as lying between Nozick's assertion of the inviolability of personal boundaries to enforceable moral obligation and the disregard of such boundaries in the utilitarian definition of the enforceable goal of morality as the greatest aggregate happiness. Unlike Nozick, and like the utilitarian, Rawls does think contribution to the well-being of others is an enforceable moral requirement; but in developing his concept of this requirement Rawls attends carefully to the boundaries of persons.⁴

Within only a single life, Rawls points out, a maximization of overall happiness like that urged for all society by the utilitarian would be appropriate, since, for example, I would be the one who emerged from the hardships of my student years into those satisfactions of my later life that were made possible by the earlier sacrifices. One person would experience both, and his life's experience would be better overall because of the hardships. But, argues Rawls, in a society that is organized on utilitarian lines, in order, that is, to produce the greatest overall happiness across distinct lives, whoever ends up worst off will never escape from his relatively poor condition into an enjoyment of the increase in happiness that resulted for others from the utilitarian maximization of overall happiness. Each person, after all, can only live his own one life. To the extent we are bothered by this factor, by the fate of an individual within his boundaries, we must attend particularly to the well-being of the worst off in society. This is egalitarianism.

Thomas Nagel in his paper "Equality" has argued that each of the three positions I have just been discussing, utilitarianism, egalitarianism and libertarianism, is an expression

³ Robert Nozick, *Anarchy, State and Utopia* (New York: Basic Books, 1974), pp.32-3.

⁴ John Rawls, *A Theory of Justice* (Cambridge, Mass.: The Belknap Press of Harvard University Press, 1971), pp. 26-7, 29-30.

of a proper moral sensitivity.⁵ They are three ways of being concerned for all, of being moral, that differ, as I would put it, in the strength of the emphasis they place on the supposed boundaries of persons. (These variations of emphasis, I might add, seem to me to reflect the already described tensions in the hypothetical perfect grasp of reality, which must somehow comprehend together, and reconcile the interests of, consciousnesses that are alienated from each other.)

Thus, says Nagel, the attempt by a utilitarian, egalitarian or libertarian to establish his one position as the single embodiment of justice would be a mistake. A society that ignored liberty and equality for the sake of utility would be distorted, as would a society that ignored utility and liberty in favour of equality, or utility and equality in favour of liberty. I agree with Nagel that our actual notion of justice seems thus to require a balance of all three values. But, as I explained earlier, I also think that there is really only one subject of consciousness. I think, therefore, that the egalitarian and libertarian concepts of justice are not tied directly to reality but to the powerful illusion of personal boundaries.

These issues about the consequentialist or deontological content of morality depend on conflicts of view or emphasis regarding the basic features of reality, the reality to which the principle of best action must be applied. The principle itself remains aloof from such conflicts. For it depends on nothing but the logic of agency.

Our Actual Moral Reasoning

Let's recall the fundamental argument of this paper: Something is really desired by me if and only if I would desire it with a perfect grasp of its nature. If I am to be doing what I really want, then, I must do that which I would want if I had a perfect grasp of everything I was doing, including a perfect grasp of what it is like to be everyone who is affected by my actions. To get as close as possible to doing what I really want, then, I must respect others in my actions. This is morality.

I contend that this view of morality represents our own actual moral reasoning. Other views, of course, characterize our actual moral thinking in other ways. But how could there be such difficulty and controversy about the nature of our own reasoning? Don't we just know how we think? A brief consideration of another problematic mode of reasoning, induction, may help illuminate this issue.

David Hume famously challenged philosophy to provide an explicit justification for induction. If we observed 100 birds of a new species on a newly discovered island and found that each of these birds was blue, we then would *feel* justified in expecting that the next such bird we observed would also be blue. (Let's call these "humbirds".) But, as Hume might have put it with regard to this example, there could be no logical necessity in the next humebird being blue. And with this in mind, Hume himself surrendered to his own challenge. He reluctantly concluded that we expect further repetitions of much repeated past experiences because of nothing but mechanical habit and that it is not possible to provide an intellectual justification for induction.

I think this is wrong. There is indeed, as Hume would insist, no necessity that the next humebird be blue; but there *is* a necessity that it be *probable* that the next humebird be blue. For it is necessarily probable that this collection of random samples has a similar proportion of blueness to that of the general population from which it has been taken.

When we observe the 100 humebirds, we think in an implicit fashion of what the general population is like. If only 100 of the humebirds in the general population had been

⁵ Thomas Nagel, "Equality", in *Mortal Questions* (Cambridge: Cambridge University Press, 1979), pp.106-27.

blue, then our seeing only blue humebirds in the first 100 random observations of them would have been extremely improbable. And it is improbable that something improbable happened. That the observations were of nothing but blue humebirds gets less improbable the more purely blue is the whole population available to our sampling. The least improbable account is that the humebirds in that whole population are generally blue. It goes along with this reasoning, of course, that it is probable that the next humebird to be observed from the same population and under the same general conditions will also be blue. And this, I contend, is the implicit thinking that rightly makes us expect that this next humebird will be blue.

In our moral thinking it is the implicit thought of the full reality of others, as possessing a weight that must shape our practical decisions if these are not to be distorted, that pulls us towards a moral judgment. Both kinds of thinking, moral and inductive, have an objective basis and justification. Neither is merely a contingent, subjective reaction to the world.

Hume also famously posed the challenge to our moral reasoning. How could the compelling character of moral values, of what “ought” to be done, be rationally derived from nothing but the facts, from what “is”? And here too he surrendered in the face of the challenge, by making the authority of morality depend on a contingent motivation, on the presence in the valuer of a disinterested benevolence or sympathy, without which, thought Hume, there would be no moral values.

But we have seen how the principle of best action, derived from necessary truths about the real desires of everyone, makes morality unconditionally compelling. And in that principle, what we “ought” to do is based completely on what “is”. Only by uncovering this principle do we finally come to see, as well as feel, the authority of morality.

Socrates in Plato’s dialogue the “Protagoras” claims that all virtues are really knowledge. As an example he uses courage, a virtue that may have seemed instead to be emotional. The coward, says Socrates, makes a mistake in judging the real importance of the objects of his choices. Consider how one’s own hand, because it is closer, will fill much more of one’s vision than will a great but distant mountain. The cowardly soldier mistakenly judges the dangers that are close to him to be greater in importance than the truly greater danger of the temporarily more distant lifetime of shame that is the consequence of his cowardice. The courageous soldier is exercising the art of measuring, says Socrates, much as the man who is rational in his visual judgments exercises the art of measuring regarding the sizes of the objects he sees, by correcting for the effects of perspective.

What I have done is to extend horizontally Socrates’ art of measuring, so that it reaches into the lives of others as well as the future. This is the dimension of morality.

Morality amounts to dealing with other conscious beings in a way that corrects for the perspective of an individual’s experience. What I have been describing in this paper is what all of us do when we practise consideration for others from the motive that it is morally right to do so.