

## WHY SHOULD I CARE ABOUT MORALITY?

By Arnold Zuboff, University College London

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 the 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 try to deal with a couple of problems people have with this view.

“Surely we can't have the perfect grasp that you claim is so desirable.”

I am not claiming either that a perfect grasp is desirable or that we can have it. What I am saying is crucially more subtle than that. I am saying that what is really desirable now, for me with my limited grasp of things, is what I would recognise to be desirable if I did grasp everything perfectly. But we can know that the perfect grasp itself would not be discovered within itself to be desirable, since it would have to include within it all actual and potential miseries. The perfect grasp would rather discover that many a limited grasp would be far more desirable than the perfect grasp as a state to be in. But only the perfect grasp could define the set of real desires that must belong to any being with any grasp.

“Many things are desirable only because my perceptions are limited. Crossword puzzles would be no fun if I already knew the answers, and there are good foods that I couldn't enjoy with too vivid an awareness of what they were or where they came from.”

Yes, and a perfect grasp would have to include a perfect appreciation, as I've already indicated, of the values of limitations on awareness. Such limitations are part of what would be recognised as desirable in a perfect grasp.

For a much fuller discussion of the view I am presenting let me refer you to my paper, “Morality as What One Really Desires”, which can be found in volume 20 (1995) of the journal *Midwest Studies in Philosophy*.

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 usual 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.

Two passages from “Morality As What One Really Desires” (*Midwest Studies in Philosophy*, volume 20)

By Arnold Zuboff, University College London

### FIRST PASSAGE

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.

## SECOND PASSAGE

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.