3

Self-Constitution in the Ethics of Plato and Kant

1. Introduction

One of the most famous sections of Hume's *Treatise* begins with these words:

> Nothing is more usual in philosophy, and even in common life, than to talk of the combat of passion and reason, to give the preference to reason, and to assert that men are only so far virtuous as they conform themselves to its dictates. Every rational creature, 'tis said, is oblig'd to regulate his actions by reason; and if any other motive or principle challenge the direction of his conduct, he ought to oppose it, 'till it be entirely subdu'd, or at least brought to a conformity with that superior principle. (T 2.3.3,413)

As Hume understands these claims, reason and passion are two forces in the soul, each a source of motives to act, and virtue consists in the person going along with reason. Why should the person do that? Hume tells us that in philosophy:

> The eternity, invariableness, and divine origin of [reason] have been display'd to the best advantage: the blindness, unconstancy, and deceitfulness of [passion] have been as strongly insisted on. (T 2.3.3,413)

Hume proposes to "shew the fallacy of all this philosophy," but in his demonstration he does not exactly deny what I will call "the Combat Model." He simply argues that reason is not a force, and therefore that there is no combat.

I think that there are a few questions Hume should have asked first, for the Combat Model makes very little sense. From the third-person perspective, we do sometimes explain a person's actions as the result of one motive being "stronger" than another, for instance when the person has conflicting passions. But is the difference between reason and passion then pretty much the same as the difference between one passion and another? And are a person's actions merely the result of the play, or rather the combat, of these forces within her? How then would actions be different from blushes or twitches or even biological processes?

Now we may try to solve this last problem by bringing the person, the agent, back into the picture—action is different from other physical movements because the person *chooses* to follow either reason or passion. But this makes the Combat Model even more perplexing. For what is the essence of this person, in whom reason and passion are both forces, *neither* of them identified with the person herself, and between which she is to choose? And if the person identifies neither with reason nor passion, then how—on what principle—can she possibly choose between them? The philosophers Hume describes here seem to be imagining that the person chooses between reason and passion by assessing their merits—reason is divine and reliable, passion blind and misleading. But surely that presupposes that the person *already* identifies with reason, which is what assesses merits. But how then could the person choose passion over reason? The Combat Model gives us no clear picture of the *person* who chooses between reason and passion.

The tradition supplies us with another model of the interaction of reason and passion in the soul, which makes better sense, because it assigns to them functional and structural differences.[1] I call it the Constitutional Model, because its clearest appearance is in Plato's *Republic*, where the human soul is compared to the constitution of a polis or city-state. I believe that the Constitutional Model has important implications for moral philosophy, and my project in this essay is to spell these implications out. Specifically, the Constitutional Model implies a certain view about what an *action* is, which in turn has implications about what makes an action good or bad. These implications are a little difficult to articulate clearly in advance of the argument, but the main idea is this: what distinguishes action from mere behavior and other physical movements is that it is *authored*—it is in a quite special way attributable to the *person* who does it, by which I mean, the *whole* person. The Constitutional Model tells us that what makes an action yours in this way is that it springs from and is in accordance with your constitution. But it also provides a standard for good action, a standard that tells us which actions are most truly a person's own, and therefore which actions are most

[1] One might think that Hume is also presenting a constitutional model, since his own argument suggests that the function of passion is to determine our ends and the function of reason is to discover means to ends. Elsewhere, however, I have argued that Hume does not really believe in a principle of instrumental *practical reason*, which instructs us to take the means to our ends, and which would be needed to integrate the two functions (the determination of the end and the identification of the means) into *a single system which produces actions*. Because of that, Hume is unable to work up a *person* out of these meager resources. What I've just said will become clearer as this essay proceeds, for it is actually, in a sense, a short version of the whole argument of this essay. For the argument that Hume does not believe in a principle of instrumental practical reason, see my "The Normativity of Instrumental Reason," Essay 1 in this volume, especially pp. 32–46.

truly *actions*. Now this is the hard part to say in advance of the argument: The actions which are most truly a person's own are precisely those actions which most fully unify her and therefore most fully constitute her as their author. They are those actions that both issue from, and give her, the kind of volitional unity that she must have if we are to attribute the action to her as a whole person. What makes an action bad, by contrast, is that it springs in part not from the person but from something at work *in* or *on* the person, something that threatens her volitional unity. I sum these claims up by saying that according to the Constitutional Model, action is self-constitution.

2. Plato

In Book 1 of the *Republic*, Socrates and his friends discuss the question what justice is. The discussion is interrupted by Thrasymachus, who asserts that the best life is the unjust life, the life lived by the strong, who impose the laws of justice on the weak, but ignore those laws themselves. The more completely unjust you are, Thrasymachus says, the better you will live, for pickpockets and thieves, who commit small injustices, get punished, while tyrants, who enslave whole cities and steal their treasuries, lead a glorious life, and are the envy of everyone (R 336b–339d). Socrates, distracted by these claims, drops the discussion of what justice is, and takes up the question whether the just or the unjust life is best.

Socrates proceeds to construct three arguments designed to show that the just life is the best. The one that is central to my own argument goes like this (R 351b–352c): Socrates asks Thrasymachus whether a band of robbers and thieves with a common unjust purpose would be able to achieve that purpose if they were unjust to each other. Thrasymachus agrees that they could not do that. Justice, as Socrates says, is what brings a sense of common purpose to a group, while injustice causes hatred and civil war, and makes the group "incapable of achieving anything as a unit" (R 352a). Thrasymachus is then induced to agree that justice and injustice have the same effect wherever they occur, and therefore, the same effect within the individual human soul as they have in a group. Injustice, therefore, makes an individual "incapable of achieving anything, because he is in a state of civil war and not of one mind." The more complete this condition is the worse it is, for according to Socrates "those who are all bad and completely unjust are completely incapable of accomplishing anything" (R 352c).[2]

[2] The other two arguments are the "outdoing" argument used to establish that justice is a form of virtue and knowledge (R 349a–350d) and the function argument used to establish that the just person is happiest (R 352d–354a).

Now there's nothing obviously wrong with this argument, except of course that it flies in the teeth of the fact that we seem to see unjust people all around us, doing and accomplishing things right and left. So what is Socrates talking about? The argument leaves Socrates's audience puzzled and dissatisfied. So Plato's brothers, Glaucon and Adeimantus, demand that Socrates return to the abandoned question, what justice is, and what effect it has in the soul. It is this demand that sets Plato off on his attempt to identify justice in a larger and more visible object, the ideal city, and his famous comparison between the constitution of the city and the constitution of the soul.

It will help to review the main elements of that comparison. Plato identifies three classes in the city. First there are the rulers, who make the laws and policies for the city, and handle its relations with other cities. Second, there are the auxiliaries, a kind of combination soldier and police force, who enforce the laws within the city and also defend it from external enemies, following the orders of the rulers. The rulers are drawn from the ranks of these auxiliaries, and the two groups together are called the guardians. And finally there are the farmers, craftspeople, merchants, and so forth, who provide for the city's needs.

The virtues of the ideal city are then identified with certain properties of and relations between these parts. The wisdom of the city rests in the wisdom of its rulers (R 428b–429a). We aren't told much about this at first, except that the rulers of the ideal city, unlike Thrasymachus's rulers, rule with a view to the good of the city as a whole, and not with a view to their own good. The courage of the city rests in the courage of its auxiliaries, which is identified with their capacity to preserve certain beliefs, instilled in them by the rulers, about what is to be feared, in the face of temptation, pleasure, pain, and fear itself (R 429a–430c). The city's *sophrosyne*—its moderation or temperance—rests in the agreement of all the classes in the city about who should rule and be ruled (R 430e–432b). And its justice rests in the fact that each class in the city does its own work, and no one tries to meddle in the work of anyone else (R 433a ff.).

Plato then undertakes to find the same three parts in the human soul. The Constitutional Model, like the Combat Model, starts off from the experience of inner conflict. Socrates puts it forth as a principle that if we find in the soul opposite attitudes or reactions to a single thing at the same time, we must suppose that the soul has parts (R 436b–c). For example, the soul of a thirsty person is impelled by its thirst towards drinking. So if the soul at the very same time draws back from drinking, it must be with a different part. And this is an experience people actually have: there are thirsty people who decide not to

drink. This happens when they judge that the drink will be bad for them. As Socrates says:

> Isn't there something in their soul, bidding them to drink, and something different, forbidding them to do so, that overrules the thing that bids? . . . Doesn't that which forbids in such cases come into play as a result of rational calculation? (R 439c–d)

So reason and appetite must be two different parts of the soul.

In fact, however, Socrates's emphasis on conflict is slightly misleading, for, even if there is no conflict, two parts of the soul may be discerned. Suppose instead that the drink has nothing wrong with it, and the person who is thirsty does drink. In this kind of case, Socrates says,

> the soul of someone who has an appetite for a thing wants what he has an appetite for and takes to himself what it is his will to have, and . . . insofar as he wishes something to be given to him, his soul, since it desires this to come about, nods assent to it as if in answer to a question. (R 437c)

The soul does not act directly from appetite, but from something that endorses the appetite and says yes to it. Even when conflict is absent, then, we can see that there are two parts of the soul.

Socrates next argues that there is a third part, *thymos* or spirit, which is distinct from both reason and appetite, although it is the natural ally of reason (R 439e–441c). That it is distinct from appetite shows up in the fact that anger and indignation, which are manifestations of spirit, are often directed against the appetites themselves. This is illustrated by the story of Leontius, who was disgusted at himself for wanting to look at some corpses, and berated his own eyes for their evil appetites (R 439e–440a). Socrates claims that spirit always fights on reason's side in a case of conflict between reason and appetite. Yet it is distinct from reason, for it is present in small children and animals, who don't have reason; and, furthermore, it sometimes needs to be controlled by reason (R 440e–441c).

By these arguments Socrates establishes that the soul has the same three parts as the city. Reason corresponds to the rulers and its function is to direct things, for the good of the whole person. Spirit corresponds to the auxiliaries and its function is to carry out the orders of reason. The appetites correspond to the rest of the citizens, and their business is to supply the person with whatever he needs.

Now if the soul has parts the question is going to arise what makes them one, what unifies them into a single soul. And part of the answer is that the parts of the soul must be unified—they *need* to be unified, like the people in a city—in order to act. Specifically, we can see the three parts of the soul

as corresponding to three parts of a deliberative action. Deliberative action begins from the fact we have certain appetites and desires. We are conscious of these, and they invite us to do certain actions or seek certain ends. Since we are rational, however, we do not act on our appetites and desires automatically, but instead decide whether to satisfy them or not. As Socrates put it in a passage we looked at a moment ago, it is as if what appetite does is put a request to reason, and reason says yes or no. And then finally there is carrying the decision out—actually doing what we have decided to do. For of course we don't always do what we have decided to do, but are sometimes distracted by pleasure or pain or fear from the course we have set for ourselves. So we can identify three parts of a deliberative action corresponding to Plato's three parts of the soul, namely:

Appetite makes a proposal.
Reason decides whether to act on it or not.
Spirit carries reason's decision out.

This line of thought supports Plato's analogy between the city and the soul. For a city also engages in deliberative actions: it is not just a place to live, but rather a kind of agent that performs actions and so has a life and a history. And we can see the same three parts in a political decision. The people of the city make a proposal: they say that there is something that they need. They ask for schools, or better health care, or more police protection. The rulers then decide whether to act on the proposal or not. They say either "yes" or "no" to the people. And then the auxiliaries carry the ruler's decisions out.

In fact, the main purpose of a literal political constitution is precisely to lay out the city's mode of deliberative action, the procedures by which its collective decisions are to be made and carried out. A constitution defines a set of roles and offices that together constitute a procedure for deliberative action, saying who shall perform each step and how it shall be done. It lays out the proper ways of making proposals (say, by petition, or the introduction of bills, or whatever), of deciding whether to act on these proposals (the legislative function), and of carrying out the resulting decisions (the executive function). And in each case it says who is allowed to carry out the procedures it has specified. The constitution in this way makes it possible for the citizens to function as a single collective agent.

And this explains Socrates's puzzling definition of justice. Justice, he says, is "doing one's own work and not meddling with what isn't one's own" (R 433a–b). When Socrates first introduces this principle into the discussion (R 369e ff.), he's talking about the specialization of labor, and that's what

the principle sounds like it's about.[3] But if we think of the constitution as laying out the procedures for deliberative action, and the roles and offices that constitute those procedures, we can see what Socrates's point is. For usurping the office of another in the constitutional procedures for collective action is *precisely* what we mean by injustice, or at least it is one thing we mean. For instance, if the constitution says that the president cannot make war without the agreement of the congress, and yet he does, he has usurped congress's role in this decision, and that's unjust. If the constitution says that each citizen gets to cast one vote in the election, and through some fraud you manage to vote more than once, you are diminishing the voice of others in the election, and that's unjust. So injustice, in one of its most familiar senses, is usurping the role of another in the deliberative procedures that define collective action. It is meddling with somebody else's work.

I said in one sense, for this is very much what is sometimes called a *procedural* conception of justice, as opposed to a *substantive* one. This distinction represents an important tension in our concept of justice, and a standing cause of confusion about the source of its normativity. On the one hand, the idea of justice essentially involves the idea of following certain procedures. In the state, as I have been saying, these are the procedures which the constitution lays out for collective deliberative action: for making laws, waging wars, trying cases, collecting taxes, distributing services, and all of the various things that a state does. According to the procedural conception of justice, an action of the state is just if and only if it is the outcome of actually following these procedures. That is a *law* which has been passed in form by a duly constituted legislature; this law is *constitutional* if (say) the supreme court says that it is; a person is *innocent* of a certain crime when he has been deemed so by a jury; someone is *the president* if he meets the legal qualifications and has been duly voted in, and so forth. These are all normative judgments—the terms *law*, *constitutional*, *innocent*, and *president* all imply the existence of certain reasons for action—and their normativity *derives from* the carrying out of the procedures that have established them.

On the other hand, however, there are certainly cases in which we have some independent idea of what outcome the procedures ought to generate. These independent ideas serve as the criteria for our more substantive judgments—in some cases, of what is just; in other cases, simply of what is right or best. And these substantive judgments can come in conflict with the actual outcomes of carrying out the procedures. Perhaps the law is unconstitutional, though the

[3] Socrates not only openly acknowledges this oddity later on, but actually suggests that the principle of the specialization of labor is "beneficial" because it is "a sort of image of justice" (R 443c).

legislature has passed it; perhaps the defendant is guilty, though the jury has set him free; perhaps the candidate elected is not the best person for the job, or even the best of those who ran, or perhaps due to the accidents of voter turnout he does not really represent the majority will. As this last example shows, the distinction between the procedurally just and the substantively just, right, or best, is a rough and ready one, and relative to the case under consideration. Who should be elected? The best person for the job, the best of those who actually run, the one preferred by the majority of the citizens, the one preferred by the majority of the registered voters, or the one elected by the majority of those who actually turn out on election day . . .? As we go down the list, the answer to the question becomes increasingly procedural; the answer above it is, relatively, more substantive. We may try to design our procedures to secure the substantively right, best, or just outcome. But—and here is the important point—according to the procedural conception of justice, the normativity of these procedures nevertheless does not spring from the efficiency, goodness, *or even the substantive justice* of the outcomes they produce. The reverse is true: it is the procedures themselves—or rather the actual carrying out of the procedures—that confers normativity on those outcomes. The person who gets elected holds the office, no matter how far he is from being the best person for the job. The jury's acquittal stands, though we later discover new evidence that after all the defendant was guilty.

Now if the normativity of the outcomes springs from the carrying out of the procedures, where, we may ask, does the normativity of the procedures themselves come from? And here we run into the cause of confusion I mentioned at the outset, for there is a standing temptation to believe that the procedures themselves must derive their normativity from the good quality of their outcomes. That cannot be right, as I've just been saying, for if the normativity of our procedures came from the substantive quality of their outcomes, we'd be prepared to set those procedures aside when we knew that their outcomes were going to be poor ones. And as I've just been saying, we don't do that. Where constitutional procedures are in place, substantive rightness, goodness, bestness, or even justice is neither necessary nor sufficient for the normative standing of their outcomes.

Perhaps we may now be tempted to say that what makes the procedures normative is the *usual* quality of their outcomes, the fact that they get it right most of the time. After all, even if we stand by the outcomes of our procedures though in this or that case they are bad, we would certainly change those procedures if their outcomes were bad *too often*. But this cannot be the whole answer, both because it isn't always true—think of the jury system—but also because, as act utilitarians have been telling us for years, it is irrational to

follow a procedure merely because it usually gets a good outcome, when you know that this time it will get a bad one. So perhaps we should say that the normativity of the procedures comes from the usual quality of their outcomes *combined* with the fact that we must have some such procedures, and we must stand by their results. But why must we have such procedures? Because without them collective action is impossible. And now we've come around to Plato's view. In order to act together—to make laws and policies, apply them, enforce them—in a way that represents, not some of us tyrannizing over others, but all of us acting as a unit, we must have a constitution that defines the procedures for collective deliberative action, and we must stand by their results.[4]

According to Plato, the normative force of the constitution *consists* in the fact that it makes it possible for the city to function as a single unified agent. For a city without justice, according to Plato, above all lacks unity—it is not one city, he says, but many (R 422d–423c; see also R 462a–e). When justice breaks down, the city falls into civil war, as the rulers, the soldiers, and the people all struggle for control. The deliberative procedures that unify the city into a single agent break down, and the city *as such* cannot act. The individual citizens and classes in it may still perform various actions, but the city cannot act as a unit.

And this applies to justice and injustice within the individual person as well. Socrates says:

> One who is just does not allow any part of himself to do the work of another part or allow the various classes within him to meddle with each other. He regulates well what is really his own and rules himself. He puts himself in order, is his own friend, and harmonizes the three parts of himself like three limiting notes in a musical scale—high, low, and middle. He binds together those parts and any others there may be in between, and from having been many things he becomes entirely one, moderate and harmonious. Only then does he act. (R 443d–e)

But if justice is what makes it possible for a person to function as a single unified agent, then injustice makes it impossible. Civil war breaks out between appetite, spirit, and reason, each trying to usurp the roles and offices of the others. The deliberative procedures that unify the soul into a single agent break down, and the person *as such* cannot act. So Socrates's argument from Book 1 turns out to be true. Desires and impulses may operate within the unjust person, as individual citizens may operate within the unjust state. But the

[4] I have also discussed these points in "Taking the Law into Our Own Hands: Kant on the Right to Revolution," Essay 8 in this volume, pp. 246–7. The discussion here is in large part lifted from that discussion.

unjust *person* is "completely incapable of accomplishing anything" (R 352c) because the unjust *person* cannot act at all.

3. Kant

Now let's turn to Kant. The best way to see that Kant is thinking in terms of the Constitutional Model is to consider the argument he uses to establish that the categorical imperative is the law of a free will (G 4:446–448). Kant argues that insofar as you are a rational being, you must act under the idea of freedom. And a free will is one that is not determined by any alien cause—by any law outside of itself. It is not, in Kant's language, "heteronomous." But Kant claims that the free will must be determined by some law or other—I will take up the argument for that in section 7—and so it must be "autonomous." That is, it must act on a law that it gives to itself. And Kant says that this means that the categorical imperative *just is* the law of a free will.

To see why, we need only consider how a free will must deliberate. So here is the free will, completely self-governing, with nothing outside of it giving it any laws. And along comes an inclination, and presents the free will with a proposal. Now inclinations, according to Kant, are grounded in what he calls "incentives," which are the features of the objects of those inclinations that make them seem attractive and eligible.[5] Suppose that the incentive is that the object is pleasant. Then inclination says: end-E would be a very pleasant thing to bring about. So how about end-E? Doesn't that seem like an end to-be-produced? Now what the will chooses is, strictly speaking, actions, so before the proposal is complete, we need to make it a proposal for action. Instrumental reasoning determines that you could produce end-E by doing act-A. So the proposal is: that you should do act-A in order to produce this very pleasant end-E.

Now if your will were heteronomous, and pleasure were a law to you, this is all you would need to know, and you would straightaway do act-A in order to produce that pleasant end-E. But since you are autonomous, pleasure is not a law to you: nothing is a law to you except what you make a law for yourself. You therefore ask yourself a different question. The proposal is that you should do act-A in order to achieve pleasant end-E. Since nothing is a law to you except what you make a law for yourself, you ask yourself whether you could take *that* to be your law. Your question is whether you can will the maxim of doing act-A in order to produce end-E as a law. Your question,

[5] For a more complete account of these ideas and Kant's moral psychology generally see the first section of my "Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind."

in other words, is whether your maxim passes the categorical imperative test. The categorical imperative is therefore the law of a free will.

Inclination presents the proposal; reason decides whether to act on it or not, and the decision takes the form of a *legislative act*. This is clearly the Constitutional Model.

4. Standards for Action

The main point of resemblance between the theories of Plato and Kant shows up, however, in their treatment of bad action. On the Combat Model, what happens when a person acts badly? The answer must be that the person is overcome by passion. But on the Combat Model we could just as well say that when a person acts well, she is overcome by reason, for the two forces seem to be on a footing. According to the Constitutional Model, on the other hand, a person acts well when she acts in accordance with her constitution. If reason overrules passion, she should act in accordance with reason, not because she identifies with reason, but because she identifies with her constitution, and it says that reason should rule.[6] So what happens when a person acts badly? Here we run into what looks, at first, like a difficulty for the Constitutional Model. It turns out to be the source of its deepest insights.

The difficulty is, of course, that according to the account of Plato I just gave, an unjust *person* cannot act at all, because an unjust person is not unified by constitutional rule. When a city is in a state of civil war, it does not act, although the various factions within it may do various things. The analogy suggests that when a soul is in a state of civil war, and the various forces within it are fighting for control, what looks to the outside world like *the person's actions* are really just the manifestations of forces at work *within* the person. So it looks at first as if *nothing exactly counts as a bad action*.

And there's an *exact* analogy to this difficulty in Kantian ethics. For a well-known problem in the *Groundwork* is that Kant appears to say that only autonomous action, that is, action governed by the categorical imperative, is really free action, while bad or heteronomous "action" is behavior *caused* by the work of desires and inclinations in us (G 4:453–55). But if this were so, then it would be hard to see how we could be held responsible for bad

[6] Julia Annas and others have pointed out to me that there is some tension between this idea and certain passages in the latter books of the *Republic* which strongly suggest that Plato's view is that we should identify with reason—most notably the passage at 588b–e in which Plato compares the three parts of the soul to a many-headed beast (appetite), a lion (spirit), and a human being (reason). I agree, but I think that the tension is within the text of the *Republic* itself, that it is part of a general tension between the conceptions of the soul in the earlier and later books.

or heteronomous action, or why we should even regard it as something we *do*. It seems more like something that happens in us or to us. This problem arises because of the argument by which Kant establishes the authority of the categorical imperative, the argument we just looked at. For that argument seems to show that action is *essentially* autonomous. Action must take place under the idea of freedom; and a free will must be autonomous. So it looks at first as if *nothing exactly counts as a bad action*.

It's important to observe that the *structure* of the problem in these two theories is exactly the same. Kant first identifies action with autonomous action, claiming that it is essential to action that it should be autonomous. He then identifies autonomous action with action governed by the categorical imperative, universalizable action. In exactly the same way, Plato first identifies action with action that emerges from constitutional procedure, claiming that it is essential to action that it should emerge from constitutional procedure. He then identifies action that emerges from constitutional procedure with just action. In other words, each argument first identifies an essential metaphysical property of action—autonomy in Kant's argument and constitutionality in Plato's—and then in turn identifies this metaphysical property with a normative property: universalizability in Kant's argument and justice in Plato's. And this is how the case for the normative requirement is made.

Furthermore, in both arguments the identification of the metaphysical property is an attempt to capture a specific feature of action, an important thing that distinguishes an action from a mere event, namely, that an action is *attributable* to the person who does it. The metaphysical property Plato and Kant are looking for is the one that makes it true that the action is not just something that happens in or to the person but rather is something that he as a person *does*. It is the property that makes the *person* the author of the action. Plato's explicit use of the Constitutional Model makes it clear that he is trying to identify this property. For we certainly do distinguish the actions we attribute to a city as such from the actions we would attribute only to some of the individuals in it. And the basis of this distinction is whether the action was the outcome of following constitutional procedures or not. If a Spartan attacks an Athenian, for instance, we do not conclude that *Sparta* is making war on Athens unless the attack was made by a soldier acting under the direction of the rulers: that is, unless it issues from Sparta's constitutional procedures. By the analogy, we will only attribute an action to a person, rather than to something in him, if it was directed by his reason, his ruling part. In a similar way, Kant thinks that what makes an action attributable to the person is that it springs from the person's autonomy or self-government. The exercise of the person's autonomy is what makes the action *his*, and so what makes it an action.

And so we get the problem. It is the essential nature of action that it has a certain metaphysical property. But in order to have that metaphysical property it must have a certain normative property. This explains why the action must meet the normative standard: *it just isn't action* if it doesn't. But it also seems as if it explains it rather too well, for it seems to imply that only good action really is action, and that there is nothing left for bad action to be.

Now rather than finding in this a reason for rejecting these arguments, I think we should see it as our main reason for embracing them. For what we have just observed is that, according to Plato and Kant, the moral standards we apply to actions are what I have elsewhere called "internal standards"—standards that a thing must meet in virtue of what it is.[7] An internal standard is one that arises from the nature of the object to which it applies, from the functional or teleological norms that make it the object that it is. Say that a house, for instance, is a habitable shelter. Then a good house is a house that has the features that enable it to serve as a habitable shelter—the corners are properly sealed, the roof is waterproof and tight, the rooms are tall enough to stand up in, and things like that. These internal standards are what make something *a good house*.

We need to distinguish here between something's being a good or bad *house* and it's being a house that happens to be a good or bad *thing* because of some external standard. The large mansion that blocks the whole neighborhood's view of the lake may be a *bad thing* for the neighborhood, but it is not therefore a *bad house*. A house that does not successfully shelter, on the other hand, is a bad house. Let me give this kind of badness a special name. An entity that does not meet its internal standards is *defective*.

The distinction between internal and external standards is important, because internal standards meet challenges to their normativity with perfect ease. Suppose you are going to build a house. Why shouldn't you build a house that blocks the whole neighborhood's view of the lake? Perhaps because it will displease the neighbors. Now *there* is a consideration that you may simply set aside, if you are selfish or tough enough to brave the neighbors' displeasure. But because it does not make sense to ask why a house should serve as a habitable shelter, it also does not make sense to ask why the corners should be sealed and the roof should be waterproof and tight. For if you fall too far short of the internal standard for houses, what you produce will simply not be a house. And this means that there's a sense in which even the most venal

[7] Or, sometimes, constitutive standards. I discuss the conception of an internal or constitutive standard in "The Normativity of Instrumental Reason," Essay 1 in this volume. See especially pp. 61–2. There I argue that the hypothetical imperative is an internal standard for acts of the will.

and shoddy builder must try to build a good house, for the simple reason that there is no other way to try to build a house. Building a good house and building a house are not different activities: for both are activities in which we must be guided by the functional or teleological norms implicit in the idea of a house. Obviously, it doesn't follow that every house is a good house. It does, however, follow that building bad houses is not a different activity from building good ones. *It is the same activity, badly done.*

Just actions in Plato, universalizable actions in Kant, are actions that are good *as* actions, the way a house that shelters successfully is good as a house. And if this is right, we should get the same conclusions. If justice and universalizability are internal standards, then they are not extraneous considerations whose normativity may be doubted. An agent cannot simply set aside the question whether his action is universalizable or just, for if he falls too far short of the internal standards for actions, what he produces will simply not be an action. In effect this means that even the most venal and shoddy agent must try to perform a good action, for the simple reason that there is no other way to try to perform an action. Performing a good action and performing an action are not different activities: for both are activities in which we must be guided by the functional or teleological norms implicit in the idea of an action. Obviously, it doesn't follow that every action is a good action. It does, however, follow that performing bad actions is not a different activity from performing good ones. *It is the same activity, badly done.*

5. Defective Action

So if we could make these claims plausible, or even intelligible, we would have an important result here: an answer to the question why our actions must meet moral standards. Unjust or non-universalizable actions would be *defective*: they would be bad *as actions*. But how can actions be defective, and still *be* actions? The Constitutional Model again provides us with the resources for an answer. For we all know that the action of a city may be formally or procedurally constitutional and yet not substantively just. Indeed, nothing is more familiar: a law duly legislated by the congress and even upheld by the supreme court may for all that be unjust. So it's not as if there's no territory at all between a perfectly just city and the complete disintegration of a civil war. A city may be governed, and yet be governed by the wrong law. And so may a soul. This, according to Plato and Kant, explains how bad action is possible.

In Kant's work this emerges most clearly in the first part of *Religion within the Limits of Reason Alone*. There we learn that a bad person is not after

all one who is pushed about, or caused to act, by desires and inclinations. Instead, a bad person is one who is governed by what Kant calls the principle of self-love, by a principle which subordinates moral considerations to those arising from inclination (REL 6:36). The person who acts on the principle of self-love *chooses* to act as inclination prompts (REL 6:32–39). Let me try to make it clear why Kant thinks that an action based on the principle of self-love is *defective*, rather than merely externally bad.

Imagine a person I'll call Harriet, who is, in any formal sense you like, an autonomous person. She has a human mind, is self-conscious, with the normal allotment of the powers of reflection. She is not a slave or an indentured servant, and we will place her—unlike the original after whom I am modeling her—in an advanced modern constitutional democracy, with the full rights of free citizenship and all her human rights legally guaranteed to her. In every formal legal and psychological sense, what Harriet does is *up to her*. Yet whenever she has to make any of the important decisions and choices of her life, the way that Harriet does that is to ask Emma what she should do, and then that's what she does.[8]

This is autonomous action and yet it is *defective* as autonomous action. Harriet is self-governed and yet she is not, for she allows herself to be governed by Emma. Harriet is heteronomous, not in the sense that her actions are caused by Emma rather than chosen by herself, but in the sense that she allows herself to be governed in her choices by a law outside of herself. It even helps my case here that Harriet does this because she is afraid to think for herself. For, as I have argued elsewhere, this is how Kant envisions the operation of the principle of self-love.[9] Kant does not envision the person who acts from self-love as actively reflecting on what he has reason to do and arriving at the conclusion that he ought to do what he wants. Instead, Kant envisions him as one who simply follows the lead of desire, without sufficient reflection. He's heteronomous, and gets his law from nature, not in the sense that it causes his actions, but in the sense that he allows himself to be governed by its suggestions—just as Harriet allows herself to be governed by Emma's.

The analogous doctrine in Plato is much more elaborate, and this is to Plato's credit. For what Kant says here is incomplete and confusing. Minimally, it seems, Kant ought to have distinguished between a wanton principle of self-love—the principle of acting on the desire of the moment—and a prudent principle of self-love, which seeks, say, the greatest satisfaction of desires over

[8] The model for my Harriet is the persuadable Harriet Smith in Jane Austen's novel *Emma*.

[9] See my "From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action," Essay 6 in this volume, especially pp. 181–5.

time.[10] Both of these characters *are* found in Plato, and others besides. In Books 8 and 9 of the *Republic*, Plato in fact distinguishes five different ways that the soul may be governed, comparing them to five different kinds of constitutions possible for a city: the good way, which is monarchy or aristocracy; and four bad ones, growing increasingly worse: timocracy, oligarchy, democracy, and, worst of all, tyranny. In the three middle cases, what makes the constitution bad is that the unity of the person who lives under it depends upon contingent circumstances.

Nearest to the aristocratic soul is the timocratic person, who, like the city he is named for, is ruled by considerations of honor. Such a person loves the outward form, the beauty, of goodness, almost as if it were goodness itself. This person goes wrong, and becomes divided against himself, in a certain kind of case—namely, the kind of case in which the right thing is something which seems dishonorable. Suppose, for instance, the timocratic person is fighting for the good of the city, but we reach a point where really surrender is the better course. The timocratic person may be so fixed on the honorableness, the beauty, the glamour if you will, of this kind of action, of fighting-for-the-good-of-the-city, that he may be unable to give up, even though it is really for the good of the city that he should do so.[11]

Next comes the oligarchic person, who appears to be ruled by prudence: he is cautious, non-luxurious, and concerned with long-term enrichment. In describing him Plato employs a distinction between necessary desires, whose satisfaction is beneficial or essential to survival, and unnecessary or luxurious desires, which are harmful and should not be indulged. The oligarchic person is attentive to the necessary desires and to money, while he represses his unnecessary desires. But he represses them because they are unprofitable, rather than because it is bad to indulge them. The result of this forceful repression, according to Socrates, is that "someone like that wouldn't be

[10] If I am right in saying that Kant sees self-love as operating unreflectively, this might seem to favor a wanton principle of self-love. Sometimes, however, it is clear that Kant has a prudent principle of self-love in mind—see for instance C2 5:35–36. While I think that the wanton principle does square better with Kant's arguments, I also think it should be possible to make the second *Critique* passages consistent with the view that those who act from self-love are unreflective. We just need to argue that there is a difference between being *reflective* and being *calculating*.

[11] Although space constraints don't allow me to spell out the idea in sufficient detail here, I am tempted to say that the problem with the timocratic person is that he is unable to deal with the contingencies that call for the application of what I have elsewhere called, following John Rawls, "non-ideal theory" (see my "The Right to Lie: Kant on Dealing with Evil," CKE essay 5, especially pp. 147–54). That is, he acts well, except in those moments when true goodness calls for concession, compromise, a less strict rule, or even—though this is rare—actions that are formally wrong. See my "Taking the Law into Our Own Hands: Kant on the Right to Revolution," Essay 8 in this volume, for a discussion of this kind of case.

entirely free from internal civil war and wouldn't be one but in some way two." This kind of prudence rules despotically over the appetitive part, like the rich ruling over a discontented working class. Should some outside force—perhaps simply a sufficient temptation—strengthen and enliven his unnecessary desires, the oligarchic person may quite literally lose control of himself. If generally the oligarchic person manages to hang together, it is because he has the sort of imitation virtue which Socrates makes fun of in the *Phaedo*, the virtue of those who are able to master some of their pleasures and fears because they are in turn mastered by others.[12] Socrates has in mind here such arguments as that you should be temperate because that way you will get more pleasure on the whole. Generally, Plato seems to think that honor and prudence are principles of choice sufficiently like true virtue to hold a soul together through most kinds of stress, although in the oligarchic person the fault lines are increasingly visible.[13]

Next in line is the democratic person, in whom the unnecessary desires are not repressed, and who as a result is a wanton. Socrates says that the democratic person:

> puts his pleasures on an equal footing . . . always surrendering rule over himself to whichever desire comes along, as if it were chosen by lot. And when that is satisfied, he surrenders the rule to another, not disdaining any but satisfying them all equally. (R 561b)

Democracy is a degenerate case of self-government, for such a person governs himself only in a minimal or formal sense, just as choosing by lot is different only in a minimal or formal sense from not choosing at all. The coherence of the democratic person's life is completely dependent on the accidental coherence of his desires. To see the problem, consider a story:

> Jeremy, a college student, settles down at his desk one evening to study for an examination. Finding himself a little too restless to concentrate, he decides to take a walk in the fresh air. His walk takes him past a nearby bookstore, where the sight of an enticing title draws him in to look at the book. Before he finds it, however, he meets his friend Neil, who invites him to join some of the other kids at the bar next door for a beer. Jeremy decides he can afford to have just one, and goes with Neil to the bar.

[12] See Plato, *Phaedo* 68d–69c.

[13] A number of people have argued that the problem described here would not arise for the rational egoist in the more ordinary modern sense, the person who seeks to maximize the satisfaction of his own interests. Indeed this is suggested by my own remarks about how imitation virtue can help hold the oligarch together, for modern egoism is much like Plato's imitation virtue. If correct, this objection would suggest that you can constitute yourself through the egoistic principle. A full response to this objection requires a full treatment of the claim that there is a coherently formulable principle of rational egoism. See my "The Myth of Egoism," Essay 2 in this volume.

While waiting for his beer, however, he finds that the noise gives him a headache, and he decides to return home without ever having the beer. He is now, however, in too much pain to study. So Jeremy doesn't study for his examination, hardly gets a walk, doesn't buy a book, and doesn't drink a beer.[14]

Of course democratic life doesn't have to be like this; it's only an accident that each of Jeremy's impulses leads him to an action that completely undercuts the satisfaction of the last one. But that's just the trouble, for it's also only an accident if this does *not* happen. The democratic person has no resources for shaping his desires to prevent this, and so he is at the mercy of accident. Like Jeremy, he may be almost completely *incapable of effective action*.

It is from the chaos resulting from this kind of life that the tyrannical soul emerges. This kind of soul is once again unified, but not under the government of reason looking to the good of the whole. According to Plato, the tyrannical soul is governed by some erotic desire (R 572d–573a), which subordinates the entire soul to its purposes, leaving the person an absolute slave to a single dominating obsession (R 571a–580a).[15]

In Plato's story, as in Kant's, bad action is action governed by a principle of choice which is not reason's own: a principle of honor (timocracy), prudence (oligarchy), wantonness (democracy), or obsession (tyranny). It is action, because it is chosen in accordance with the exercise of a principle by which the agent rules himself and under whose rule he is—in a sense—unified. Yet it is defective, because it is not reason's own principle, and the unity that it produces is, at least in the three middle cases, contingent and unstable. And Plato can say with Kant that the person who governs himself in one of these ways isn't after all completely self-governed. For he is propped up, so to speak, by the fact that the circumstances that would create civil war in his soul don't happen to occur.

6. Good Action and the Unity of the Platonic Soul

Now we are almost ready to talk about what makes action good. But first I want to take up a possible objection. I've just said that in the conditions of

[14] I have lifted this example from footnote 52 of my "The Normativity of Instrumental Reason," essay 1 in this volume.

[15] The problem with tyranny is not the same as that with timocracy, oligarchy, and democracy—it is not that the unity it produces in the soul is contingent. Plato envisions tyranny as a kind of madness (see R 573c ff.). As I imagine the tyrant, his relation to his obsession is like a psychotic's relation to his delusion: he is able, and prepared, to organize everything else around it, but at the expense of a loss of his grip on reality, on the world. But that is only a sketch, and a fuller treatment of this principle, and of the question why a person cannot successful integrate himself under its governance, is required for the completeness of the argument of this essay.

timocracy, oligarchy, and democracy, your unity and so your self-government are propped by external circumstances, by the absence of the conditions under which you would fall apart. But what, you might ask, is so bad about that? The defect in these characters is like a geological fault line, a potential for disintegration that does not necessarily show up, and so long as it doesn't, these people have constitutional procedures and so they can act. So why not just go ahead and be, say, oligarchical? You'll hold together most of the time, you'll be able to perform actions, and you'll save all that money besides.

There is yet another way to ask this same question, which is to ask whether Glaucon's challenge is not too extreme. Glaucon wants Socrates to tell him what justice and injustice do to the soul. So he sets up the following challenge: take on the one hand a person who has a completely unjust soul, and give him all of the outward benefits of justice, that is, all the benefits that come from people believing you are just. And take on the other hand a person who has a completely just soul, and give him all of the outward disadvantages of injustice, all the disadvantages that come from people believing you are unjust (R 360d–361e). In particular, the just person who is believed to be unjust will be—and I'm quoting now—"whipped, stretched on a rack, chained, [and] blinded with fire" (R 361e). Socrates is supposed to show that it is better to be just than unjust *even then*. But isn't that too much to ask?

In the context of the argument of the *Republic*, it is not. For the question of the *Republic* is asked as a *practical* question: it is the question whether the just life is more worthy of *choice* than the unjust life. And if you choose to be a just person, and to live a just life, you are thereby choosing to do the just thing even if it means you will be whipped, stretched on the rack, chained, and blinded with fire. You can't make a conditional commitment to justice, a commitment to be just unless the going gets rough. Your justice rests in the nature of your commitments, and a commitment like that would not *be* a commitment to justice. So when deciding whether to be a just person, you've got to be convinced in advance that it'll be worth it even if things do turn out this way.

Suppose—for it's plausible enough—there's a person who lives a just life, is decent and upstanding, always does his share, never takes an unfair advantage, sticks to his word—all of that—but then, one day, he is put on the rack, and under stress of torture does something unjust. Say he divulges a military secret, or the whereabouts of a fugitive unjustly pursued. Am I saying that this shows that he was never really committed to justice, because his commitment must have been conditional? *Of course not.* What the case shows is that the range of things people can *be* is wider than the range of things they can choose, so to speak, *in advance* to be. This person was committed to keeping his secrets

on the rack, but he failed, that's all—and very understandably too. But the fact that you can be a just person who in these circumstances will fail does not show that you can decide in advance to be a just person who in these circumstances will fail: that is, it doesn't show that you can make a conditional commitment to justice. For suppose you surprise yourself and you do hold out and you keep the secret even when they put you on the rack. Did you then fail to *keep* your conditional commitment?

So Glaucon's challenge is a fair one. But Plato more than meets it. For he doesn't merely prove that the just life is the one most worthy of choice. He proves the just life is the only one you can choose. Let me try to explain why.

Consider Plato's account of the principle of just or aristocratic action. Plato says of the aristocratic soul that:

> when he does anything, whether acquiring wealth, taking care of his body, engaging in politics, or in private contracts—in all of these, he believes that the action is just and fine that preserves this inner harmony and helps achieve it, and calls it so, and regards as wisdom the knowledge that oversees such actions. And he believes that the action that destroys this harmony is unjust, and calls it so, and regards the belief that oversees it as ignorance. (R 443e–444a)

In other words, the principle of justice directs us to perform those actions that establish and maintain our volitional unity. Now we have already seen that according to Plato volitional unity is essential if you are to act as a person, as a single unified agent. So reason's own principle *just is* the principle of acting in a way that constitutes you into a single unified agent. Deliberative action is self-constitution.

In fact, deliberative action by its very nature imposes constitutional order on the soul. When you deliberate about what to do and then do it, what you are doing is organizing your appetite, reason, and spirit, into the unified system that yields an action that can be attributed to you as a person. Deliberative action pulls the parts of the soul together into a unified system. Whatever else you are doing when you choose a deliberative action, you are also unifying yourself into a person. This means that Plato's principle of justice, reason's own principle, is the *formal* principle of deliberative action. It is as if Glaucon asked: what condition could this be, that enables the just person to stick to his principles even on the rack? And it is as if Plato replied: don't look for some *further* condition which has that as an *effect*. Justice is not some other or further condition that enables us to maintain our unity as agents. It is that very condition itself—the condition of being able to maintain our unity as agents.

To see that this is formal, consider the following comparison. One might ask Kant: what principle could this be, that enables the free person to be

autonomous, to rule herself? And Kant would reply: don't look for some *further* principle that has that as an *effect*. The categorical imperative is not some other or further principle that enables us to rule ourselves. It is that very principle itself, the principle of giving laws to ourselves.

On the one hand, this account of the aristocratic soul shows us why the demands of Platonic justice are so high. On certain occasions, the people with the other constitutions fall apart. For the truly just person, the aristocratic soul, there are no such occasions. She is entirely self-governed, so that all of her actions, in every circumstance of her life, are really and fully her own: never merely the manifestations of forces at work in her or on her, but always the expression of her own choice. She is completely self-possessed: not necessarily happy on the rack—but *herself* on the rack, herself even there.

And yet at the same time, Plato's argument shows that this aristocratic constitution is the only one you can choose. For you can't, in the moment of deliberative action, choose to be something less than a single unified agent. And that means you can't exactly choose to act on any principle other than the principle of justice. Timocratic, oligarchic, and democratic souls disintegrate under certain conditions, so deciding to be one would be like making a conditional commitment to your own unity, to your own personhood. And that's not possible. For consider what happens when the conditions that cause disintegration in these constitutions actually occur. If you don't fall apart, have you failed to keep your commitment, like the conditionally just person who holds out on the rack after all? But if you do fall apart, *who is it* that has kept the commitment? If you fall apart, there is no person left. You can be a timocratic, oligarchic, or democratic person, in the same way that you can be a just person who fails on the rack. But you cannot decide in advance that this is what you will be.

Of course this doesn't mean that everyone chooses the just life. What it means is that choosing an unjust life is not a different activity from choosing a just one. It is the same activity—the activity of self-constitution—badly done.

7. Good Action and the Unity of the Kantian Will

It remains to show that this is also Kant's view; and for that we need to revisit the argument by which Kant establishes that action must be in accordance with the categorical imperative, and fill in its missing step. Kant argues that insofar as you are a rational being, you must act under the idea of freedom—and this means that you do not think of yourself, or experience yourself, as being impelled into action, but rather as deciding what to do. You take *yourself*, rather

than the incentive on which you choose to act, to be the *cause* of your action.[16] And Kant thinks that in order for this to be so, you must act on a universal law. You cannot regard yourself as the *cause* of your action—you cannot regard the action as the product of your will—unless you will universally.

To see why, let us consider what happens if we try to deny it.[17] If our reasons did not have to be universal then they could be completely particular—it would be possible to have a reason that applies only to the case before you, and has no implications for any other case. Willing to act on a reason of this kind would be what I will call "particularistic willing." If particularistic willing is impossible, then it follows that willing must be universal—that is, a maxim, in order to be willed at all, must be willed as a universal law.

Now there are two things to notice here. First of all, the question is not whether we can will a new maxim for each new occasion. We may very well do that, for every occasion may have relevant differences from the one we last encountered. Any difference in the situation that is actually relevant to the decision properly belongs in our maxim, and this means that our maxim may be quite specific to the situation at hand. The argument here is not supposed to show that reasons are general. It is supposed to show us that reasons are universal, and universality is quite compatible—indeed it requires—a high degree of specificity.

The second point is that it will be enough for the argument if the principle that is willed be willed, as I will call it, as provisionally universal. To explain what I mean by that I will use a pair of contrasts. There are three different ways in which we can take our principles to range over a variety of cases, and it is important to keep them distinct. We treat a principle as *general* when we think it applies to a wide range of similar cases. We treat a principle as universal, or, as I will sometimes say, *absolutely universal*, when we think it applies to absolutely every case of a certain sort, but all the cases must be exactly of that sort. We treat a principle as *provisionally universal* when we think it applies to every case of a certain sort, unless there is some good reason why not. The difference between regarding a principle as universal, and regarding it as provisionally universal, is marginal. Treating a principle as only provisionally universal amounts to making a mental acknowledgment, to the effect that you

[16] To put it somewhat more strictly, you take yourself to be the cause of your intelligible movements, since it is only really an *action* if you are, or to the extent that you are, the cause. I think that there are important philosophical questions, yet to be worked out, about exactly how this point should be phrased, but for now I leave the more familiar formulation in the text. I am indebted to Sophia Reibetanz and Tamar Schapiro for discussions of these points.

[17] The argument that follows made its first appearance in section 1 of the Reply in SN pp. 225–33. Another version is found in my book *Self-Constitution: Agency, Identity, and Integrity*. This essay is in fact a very short version of the argument of that book.

might not have thought of everything needed to make the principle universal, and therefore might not have specified it completely. Treating principles as general, and treating them as provisionally universal, are superficially similar, because in both cases we admit that there might be exceptions. But in fact they are deeply and essentially different, and this shows up in what happens when we encounter the exceptions. If we think of a principle as merely general, and we encounter an exception, nothing happens. The principle was only general, and we expected there to be some exceptions. But if a principle was provisionally universal, and we encounter an exceptional case, we must now go back and revise it, bringing it a little closer to the absolute universality to which provisional universality essentially aspires.

The rough causal principles with which we operate in everyday life (I am not talking now about quantum physics) are provisionally universal, and we signal this sometimes by using the phrase "all else equal." The principle that striking a match causes a flame holds all else equal, where the things that have to be equal are that there is no gust of wind or splash of water or oddity in the chemical composition of the atmosphere that would interfere with the usual connection. There are background conditions for the operation of these laws, and without listing and possibly without knowing them all, we mention that they must be in place when we say "all else equal." Although there are certainly exceptions, natural law is not merely general, for whenever an exception occurs, we look for an explanation. Something must have made this case different: one of its background conditions was not met.

To see how it works in the practical case, consider a standard puzzle case for Kant's universalizability criterion. It may seem as if wanting to be a doctor is an adequate reason for becoming a doctor, for there's nothing wrong with being a doctor—in fact, really, it's rather admirable—and if you ask yourself if it could be a law that everyone who wants to be a doctor should become one, it seems, superficially, fine. But then the objector comes along and says, but look, suppose *everyone* actually wanted to be a doctor and nobody wanted to be anything else. The whole economic system would go to pieces, and then you couldn't be a doctor, so your maxim would have contradicted itself! So does this show that it is wrong to be a doctor simply because you want to?

What it shows is that the mere desire to enter a certain profession is only a provisionally universal reason for doing so. There's a background condition for the rightness of being a doctor because you want to, which is that society has some need for people to enter this profession. In effect the case does show that it's wrong to be a doctor merely because you want to—the maxim needs revision, for it is not absolutely universal unless it mentions as part of your

reason for becoming a doctor that there is a social need. Someone who decides to become a doctor in the full light of reflection also takes that into account.

That case is easy, but there's no general reason to suppose we can think of everything in advance. When we adopt a maxim as a universal law, we know that there might be cases, cases we haven't thought of, which would show us that it is not universal after all. In that sense we can allow for exceptions. But so long as the commitment to revise in the face of exceptions is in place, the maxim is not merely general. It is provisionally universal.

So particularistic willing is neither a matter of willing a new maxim for each occasion, nor is it a matter of willing a maxim that you might have to change on another occasion. Both of those are compatible with regarding reasons as universal. Instead, particularistic willing would be a matter of willing a maxim for exactly this occasion without taking it to have any other implications of any kind for any other occasion. You will a maxim thinking that you can use it just this once and then so to speak discard it; you don't even need a reason to change your mind.

identification & agency

Now I'm going to argue that that sort of willing is impossible. The first step is this: to conceive of yourself as the cause of your actions is to identify with the principle of choice on which you act. A rational will is a self-conscious causality, and a self-conscious causality is aware of itself as a cause. To be aware of yourself as a cause is to identify yourself with something in the scenario that gives rise to the action, and this must be the principle of choice. For instance, suppose you experience a conflict of desire: you have a desire to do both A and B, and they are incompatible. You have some principle that favors A over B, so you exercise this principle, and you choose to do A. In this kind of case, you do not regard yourself as a mere passive spectator to the battle between A and B. You regard the choice as yours, as the product of your own activity, because you regard the principle of choice as expressive, or representative, of yourself. You must do so, for the only alternative to identifying with the principle of choice is regarding the principle of choice as some third thing in you, another force on a par with the incentives to do A and to do B, which happened to throw in its weight in favor of A, in a battle at which you were, after all, a mere passive spectator. But then you are not the cause of the action. Self-conscious or rational agency, then, requires identification with the principle of choice on which you act.

The second step is to see that particularistic willing makes it impossible for you to distinguish yourself, your principle of choice, from the various incentives on which you act. According to Kant, you must always act on some incentive or other, for every action, even action from duty, involves a decision on a proposal: something must suggest the action to you. And in order to will

particularistically, you must in each case wholly identify with the incentive of your action. That incentive would be, for the moment, your law, the law that defines your agency or your will.

It's important to see that if you had a particularistic will you would not identify with the incentive as representative of any sort of type, since if you took it as a representative of a type you would be taking it as universal. For instance, you couldn't say that you decided to act on the inclination of the moment, *because you were so inclined*. Someone who takes "I shall do the things I am inclined to do, whatever they might be" as his maxim has adopted a universal principle, not a particular one: he has the principle of treating his inclinations *as such* as reasons. A truly particularistic will must embrace the incentive in its full particularity: it, in no way that is further describable, is the law of such a will. So someone who engages in particularistic willing does not even have a democratic soul. There is only the tyranny of the moment: the complete domination of the agent by something inside him.

Particularistic willing eradicates the distinction between a person and the incentives on which he acts. But then there is nothing left here that is the *person*, the agent, that is his will as distinct from the play of incentives within him. He is not one person, but a series, a mere conglomeration, of unrelated impulses. There is no difference between someone who has a particularistic will and someone who has no will at all. Particularistic willing lacks a subject, a person who is the cause of these actions. So particularistic willing isn't willing at all.

If a particularistic will is impossible, then when you will a maxim you must take it to be universal. If you do not, you are not operating as a self-conscious cause, and then you are not willing. To put the point in familiar Kantian terms, we can only attach the "I will" to our choices if we will our maxims as universal laws.[18] The categorical imperative is an internal standard for actions, because conformity to it is constitutive of an exercise of the will, of an action of a person as opposed to an action of something within him.

And this argument also shows that Kant's view is the same as Plato's. For if particularistic willing is what breaks us down, universal willing is what holds us together. For Kant, as for Plato, deliberative action by its very nature imposes unity on the will. It is only when you ask whether your maxim can be a universal law that you exercise the self-conscious causality, the autonomy, that yields an action that can be attributed to you as a whole person. So whatever else you are doing when you choose a deliberative action, you are also unifying yourself into a person. For Kant, as for Plato, action is self-constitution.

[18] I owe this formulation of my point to Govert den Hartogh.

8. Conclusion

I will conclude by reviewing the course of the argument and saying what I take it to have established. I started by criticizing the Combat Model for failing to identify the person who is the author of her actions. I hope that by now it is clear why it fails in that way. The Combat Model is not a picture of the human soul. It is a picture of the human soul in ruins, torn apart by civil war and therefore unable to act. According to the Constitutional Model, an action is yours when it is chosen in accordance with your constitution. Your constitution is what gives you the kind of volitional unity you need to be the author of your actions. And it is the person who acts in accordance with the best constitution, the most unified constitution, who is most truly the author of her actions. For Kant as for Plato, integrity is the metaphysical essence of morality.

The argument of this essay does not, by itself, get us all the way to the necessity of acting morally. The aim of the argument has been to establish that the Platonic principle of justice and Kant's categorical imperative are formal standards of deliberative action. Both Kant and Plato believed that a certain content, the content of ordinary morality, could be derived from these formal principles. Plato's conviction appears at one of the most notorious moments of the *Republic*, when Socrates proposes to Glaucon that they can dispel any doubts they might have about their definition of justice "by appealing to ordinary cases" (R 442d–e). Accordingly, he asks Glaucon whether the just person as Socrates has described him would embezzle deposits, rob temples, steal, betray his friends or his city, violate his oaths or his other agreements, commit adultery, be disrespectful to his parents or neglect the gods, to all of which Glaucon says, with a complaisance startling to the reader, no, he would not, the just person Socrates has described him would not do these kinds of things. We are not told exactly why he is so sure. Kant, of course, does try to show us how content can be derived from his formal principle, and to that extent his version of the argument is superior to Plato's, although the success of his efforts is the subject of an old and famous debate. I think Kant's case can be made, but I haven't been trying to do that here.[19] Both Plato and Kant's arguments move (1) from the metaphysical property of action that makes it authored and so makes it action—autonomy in Kant's case, constitutionality in Plato's—to (2) a formal normative requirement that actions must meet

[19] In lecture 3 of *The Sources of Normativity* I give an argument that aims to move from the formal version of the categorical imperative to moral requirements by way of Kant's Formula of Humanity. See especially SN 3.3.7–3.4.10, pp. 112–25.

if they are to have that property—universalizability in Kant's case, justice in Plato's—and then through (3) the derivation of content from the formal requirement to arrive at ordinary moral requirements. It is the first two steps that have been my subject here.

At least in the formal sense, then, Platonic justice and Kant's categorical imperative are internal standard for actions, because it is only insofar as your actions issue from your whole person, rather than something in you, that they can be actions. It doesn't exactly follow that we ought to choose actions justly and in accordance with the categorical imperative, for in a sense we cannot possibly choose in any other way. Choosing bad actions is not a different activity from choosing good ones. It is the same activity—the activity of self-constitution—badly done.[20]

[20] I have discussed this essay or the longer unpublished manuscript from which it is drawn with audiences at the inaugural meeting of the Society for Ethics, the University of Amsterdam, the University of Constance, the Humboldt University of Berlin, the University of Pittsburgh, the University of Virginia, the University of Salzburg, the University of Toronto, York University of Toronto, and the University of Zurich. I am grateful to all of these audiences for their interest, insightful comments and challenging questions. I would like to thank Charlotte Brown, Barbara Herman, Govert den Hartogh, Anton Leist, Richard Moran, Amélie Rorty, and Theo van Willigenburg for reading and commenting on the manuscript.