

# Self-Constitution: Action, Identity, and Integrity

## Lecture Five

### The Constitutional Model, and Bad Action

Christine M. Korsgaard

5.1.1 In my last lecture, I argued that self-consciousness produces the parts of the soul. It requires us to substitute principles of reason for our instincts, and it transforms our incentives into inclinations. At the same time, it makes it necessary for us to deliberate about what we are going to do. Since actions must be assignable to the person as a whole, the work of practical deliberation, the work that leads to action, is also a kind of reunification.

5.1.2 But there are different views about how this happens. As Hume points out – and I'm quoting now:

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 ... virtuous [when] they conform themselves to its dictates. Every rational creature, [it is] 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, [until] it be entirely subdu'd, or at least brought to a conformity with that superior principle.

As Hume understands these claims, reason and passion are two forces in the soul, each of them a source of motives to act, and virtue consists in the person following the dictate of 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: [while] the blindness, [inconstancy], and deceitfulness of [passion] have been as strongly insisted on.

Hume proposes to show us “the fallacy of all this philosophy,” but in his demonstration he does not exactly deny what I am going to call “the Combat Model” of the soul. He simply argues that reason is not after all a force, and therefore that there is no combat.

According to the Combat Model, the difference between reason and passion is pretty much the same as the difference between one passion and another: they are two forces, each urging a certain action upon the soul. Deliberative unification takes place when one side wins. There are actually two versions of the Combat Model, but neither of them can make any sense of action. According to the first version, the person's actions are just the result of the play, or rather of the combat, of these forces within her. But as I have observed before, action cannot just be the result of forces working in or on an agent. If the movement is to be assignable to the agent in the way that the idea of action requires, then the agent must be something over and above the forces working on her, something that can intelligibly be said to determine herself to action.

Now it may seem as if the obvious way to solve this problem is to bring the person, the agent, back into the picture, and to say that she chooses between reason and passion. And Hume's description suggests this – the person, he says, “gives the preference” to reason. But this second version of the Combat Model is even more perplexing than the first. 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, since that is the part of us that assesses merits. How then could the person ever choose passion over reason? The Combat Model does not enable us to form any picture of the agent who chooses between reason and passion. And this is not surprising, for on the first version of the combat model there is no agent, while the second version presupposes an agent who can have no essence and who must always already exist.

5.1.3 The tradition supplies us with another model of the interaction of reason and passion in the soul, which makes better sense, because it assigns them functional and structural differences. 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. The constitutional model, unlike the first version of the combat model, conceives the agent as something over and above her parts. But the agent is not, as in the second version of the combat model, a separately existing entity who chooses to identify with one of those parts. Instead, the agent is something over and above her parts the way the constitution of a city is something over and above the citizens and officials who live there. If the agent conforms to the dictate of reason, it is not because she identifies with reason, but rather because she identifies with her constitution, and it says that reason should rule. Following Plato, I will argue that this model can explain action - it can explain how a movement can be attributed to an agent as a unified whole. Like Kant's theory, Plato's view has the consequence that goodness is a constitutive standard for action. And yet we can also use it to explain how bad action is possible, as I will show.

5.2.1 Let me start by reminding you how the model gets on the table. In Book I 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 (336b-334d). 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.

5.2.2 Socrates proceeds to construct three arguments designed to show that the just life is best. The one that is important to us goes like this (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” (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 as Socrates tells us “those who are all bad and completely unjust are completely incapable of accomplishing anything” (352c).

5.2.3 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 can Socrates be talking about? The argument leaves his audience puzzled and dissatisfied. So Plato's brothers, Glaucon and Adiemantus, 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 city and the soul.

5.2.4 Let me 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 (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 just for 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, which are instilled in them by the rulers, about what is to be feared, in the face of temptation, pleasure, pain,

and fear itself (429a-430c). The auxiliaries are able to hold onto their belief, for instance, that nothing is more to be feared than the loss of the city's freedom, even in the face of danger to themselves. 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 (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 (433a ff.).

Plato then 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.

5.2.5 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 self-conscious, however, we do not act on our appetites and desires automatically, but instead decide whether to satisfy them or not. 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 highways, 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. And it is only when this happens, when these procedures are followed, that we attribute the action to the city. 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. According to the analogy, we will only attribute an action to a person, rather than to something in him, if it was the result of his reason acting on a proposal from his inclination – or, to put it in my earlier terms, if it involved both an incentive and a principle.

5.2.6 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 (that's

the legislative function), and of carrying the resulting decisions out (the executive function). And it says who is supposed to carry out the various steps in the procedures it has specified. The constitution in this way makes it possible for a group of 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" (433a-b). When Socrates first introduces this principle into the discussion (369eff.), he's talking about the specialization of labor, and that's what the principle sounds like it's about. 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, then he has usurped the 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.

5.2.7 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 down 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 and correctly 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, although it has been passed by the legislature or even upheld by the supreme court; 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 person who best represents the general will,

the person who comes closest to this of those who actually run, the person preferred by the majority of the citizens, the person preferred by the majority of the registered voters, the person actually elected by the majority of those who turn out on election day... As we go down that 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 goodness, rightness, *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 true representative of the general will. The jury's acquittal stands, though we later come to believe 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? Why must we follow them? 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, since if the normativity of our procedures came from the substantive quality of their outcomes, then 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 also just been saying, we don't do that. Where constitutional procedures are in place, substantive rightness, goodness, or even justice is neither necessary nor sufficient for the

normative standing of their outcomes: all that is necessary is that the procedures have actually been followed.

Perhaps you 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 do 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, not only 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 instead that the normativity of the procedures comes from the usual quality of their outcomes *combined* with the fact that we must have we must have some such procedures, and we must stand by their results. But *why* must we have some 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 together as a unit - we must have a constitution that defines the procedures for collective deliberative action, and we must stand by its results.

5.2.8 So 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 (422d-423c; see also 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 within 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 (and I've put this on the handout):

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.* (443d-e; my emphasis)

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

5.3.1 Let's go back to Kant for a moment. One of the prevailing misconceptions about Kant is that he espouses the Combat Model of the soul. To see that Kant holds the

constitutional model, we need only consider the argument he uses in the third section of the *Groundwork* to establish that the categorical imperative is the law of a rational 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 – not determined by any law that it does not choose for itself. If you have a free will then you are not, as Kant puts it, heteronomous. But Kant claims that the actions of a free will must be determined by some law or other. We have already looked at the argument for this in lecture two, the argument against particularistic willing, which shows that the will must always determine itself in accordance with some universal law. Since if you have a free will you cannot be heteronomous, and yet you must have a law, then you must be autonomous – you must act on a law that you legislate for itself. And Kant says that this means that insofar as you are rational the categorical imperative *just is* the law of your will.

To see why, we need only consider how a person with a free will must deliberate. So here you are with your free will, completely self-governing, with nothing outside of you giving you any laws. And along comes an incentive, let us say, a representation of a certain object as pleasant. Being aware of the workings of that incentive upon you, you form an inclination for the object. And that inclination takes the form of a proposal. So the inclination says: end-E would be very pleasant. So how about end-E? Doesn't that seem like an end worth pursuing? 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 universal law. Your question, in other words, is whether your maxim passes the categorical imperative test. The categorical imperative is therefore the law of a rational 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.

5.3.2 So it isn't surprising that Plato's argument seems to leave us with the very same problem that Kant's does. The Kantian imperatives, I've been claiming, are constitutive standards for action – someone who doesn't at least try to conform to them isn't acting at all. How then is bad action possible? Justice, Plato claims, is a constitutive standard for action, for someone whose movements are not determined by his constitution isn't acting at all. How then is bad action possible? And the problem arises in both cases for exactly the same reason. Conformity to the categorical imperative, in Kant's argument, and constitutional justice, in Plato's are the properties in virtue of which a movement is attributable to the person moving, and so in virtue of which it counts as his action. How then can we attribute bad action to a person, as opposed to something at work within him?

5.4.1 But the Constitutional Model also 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 substantively 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.

5.4.2. 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 his desires and inclinations. Instead, a bad person is one who is governed by what Kant calls the principle of self-love. The person who acts on the principle of self-love *chooses* to act as inclination prompts (R 6: 32-39): he takes his inclinations, without further reflection, to be reasons for action. Why is this the wrong law? Recall that if the standard of goodness is a constitutive standard, the wrong law must not be wrong in a merely external sense. The wrong law must be one that fails constitute the person's agency; the action must be, in the technical sense I introduced in Lecture one, defective. Let me try to make it clear why Kant thinks that an action based on the principle of self-love is *defective* as an action, rather than merely by some external standard bad.

According to Kant, action must be autonomous. Now imagine a person I'll call Harriet, who is, in any *formal* sense you like, an autonomous person. She has a human mind, she 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 a well-ordered modern constitutional democracy, with the full rights of free citizenship and all of her human rights legally guaranteed to her. In every *formal* legal and psychological sense we can think of, 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 try to figure out what Emma thinks she should do, and then she does that.

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 - by Emma's will. It even helps my case here that the original 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. 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 his inclinations, 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 without much thought by its proposals - just as Harriet allows herself to be governed by Emma's.

5.4.3 The analogous doctrine in Plato is much more elaborate, and this is to Plato's credit. For what Kant says here seems incomplete and confusing. Minimally, we might think, 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 time. Versions of both of these characters *are* found in Plato, and others besides. In Books VIII and IX of the *Republic*, Plato in fact distinguishes five different ways in which the soul may be governed, comparing them to five different kinds of constitutions possible for a city: the good way, which he calls monarchy or aristocracy; and four bad ones, growing increasingly worse: timocracy, oligarchy, democracy, and worst of all, tyranny. In each of these cases some part of the soul other than reason takes over the work of reason, establishing a principle that is really for its own good rather than for the good of the whole. In what follows I'll take a look at each of Plato's bad constitutions, explain what I think he has in mind, and why they are supposed to be *defective* and not just externally bad. Since the aim of the constitution is to unify the soul, the defective constitutions must lead to disunity and to that extent undercut agency. The good constitution - the aristocratic soul - will by contrast be truly unified. That constitution will be my topic next week.

5.4.4 Nearest to aristocratic soul is the timocratic person, who, like the city he is named for, is ruled by the spirited part of his soul: by the sense of honor and the love of victory. Recall that the function of spirit, according to Plato, is to preserve a belief laid down by the rational part of the soul about what is to be feared - say, for instance, that nothing is more to be feared than the loss of the city's freedom. Now take a character like this: He says that he's fighting for the freedom of the city, but if he keeps on with the battle at this rate, there won't be any city left to be free. The buildings are all in ruins and the stores have all been looted and there are so many wounded citizens that we can't take care of them all. And we begin to suspect that he doesn't exactly care about the freedom of the

city, not really, but rather that the idea of fighting-for-the-freedom-of-the-city has a certain aesthetic character, a kind of moral glamour if you will, and he's got fixed on that, and become quite heedless of the what actually is happening to the city. This is a person in whom spirit, the sense of honor, has usurped the role of reason. Most of the time, of course, the person ruled by honor does better, for he loves the outward manifestation, the beauty of goodness, just as if it were goodness itself. Indeed he cannot distinguish the two, and that is his problem: the work of spirit is to preserve a belief, not to reflect on it. Spirit is a source of incentives, and it preserves the belief by building it into the person's representation of the world; giving up the fight is dishonorable and so it *looks* wrong to him, and that's why he won't do it. I am tempted to say that the problem with the timocratic person is that he is unable to deal with those contingencies that call for the application of what I have elsewhere called, following John Rawls, "non-ideal theory." That is, to put the point roughly, he does fine, except in those moments when what the situation actually calls for is concession, compromise, a bending of the rules, or even – say as in a case of civil disobedience – actions that are in some formal sense wrong. So in this kind of case, while fighting for the freedom of the city he destroys the city; in this kind of case, although perhaps only here, an incoherence in his will makes its appearance, destroying his efficacy and his agency with it.

5.4.5 Next comes the oligarchic person, who in Plato's account appears to be ruled by prudence, in a sense that's somewhere between the contemporary philosophical sense – someone who tries to maximize his own satisfaction – and the more everyday sense of being cautious, non-luxurious, and concerned with long-term enrichment. Now for want of time, I'm not going to talk about Plato's prudent person today. Instead I'm going to go

directly to his modern descendent – the contemporary rational egoist who, we are told, aims to maximize the satisfaction of his desires. For this character seems to many people to be the primary rival to the good person. After all, he has a way of organizing his inclinations – namely maximization – into a unified goal. So it seems to follow that he will also have a unified will.

Well, first we must make sure we know who he is. The view that we are to maximize the satisfaction of our desires is ambiguous, because the idea of “satisfaction” is ambiguous. “Satisfaction” may refer either to an objective or a subjective state. Objective satisfaction is achieved when the state of affairs that you desire is in fact realized. For instance, you want your painting to hang in the Metropolitan Museum of Art, and it does. Obviously, you could achieve the satisfaction of your desire in the objective sense without knowing anything about it: you may never know that your dream of artistic fame has been realized. Subjective satisfaction by contrast is a sort of pleasurable consciousness that objective satisfaction obtains. You know that your picture has been hung in the Museum, say, and you feel good about it; you reflect on the fact with pleasure. Although subjective satisfaction is pleasurable, it is important to distinguish it from pleasure in general. Rational egoism is not supposed to be the same thing as hedonism. Subjective satisfaction is a specific kind of pleasure, pleasure taken in the knowledge or belief that a desire has been satisfied. Is this what the rational egoist tries to maximize?

Someone who deliberates with the aim of achieving the maximum sense of subjective satisfaction over the whole course of his life does seem to be in a recognizable sense egoistic. His conduct is governed by the pursuit of something that will be experienced as a good by himself. But there is a problem about saying that he is rational. Subjective satisfaction is the pleased perception of objective satisfaction and so is

conceptually dependent upon objective satisfaction. And so, one would think, its importance must be dependent on the importance of objective satisfaction as well. There would be something upside down about thinking it mattered that you should achieve subjective satisfaction independently of thinking that it mattered that you should achieve objective satisfaction. You can see the problem by imagining a case in which they pull apart. John Rawls used to tell the following story in his classes.

A man is going away to fight in a war, in which he may possibly die. The night before he leaves, the devil comes and offers him a choice. Either while he is away, his family will thrive and flourish, but he will get word that they are suffering and miserable; or while he is away his family will suffer and be miserable, but he will get word they are thriving and happy. He must choose now, and of course he will be made to forget that his conversation with the devil and the choice it resulted in ever took place.

The problem is obvious. The man loves his family and wants them to be thriving and happy, and this clearly dictates the first choice, where his family thrives but he believes they do not. But the goal of achieving subjective satisfaction seems to favor the second choice, where he gets to enjoy the satisfaction of believing they thrive when actually they do not. So here we have *rationality* supposedly dictating the choice of a pleasing delusion over a state of affairs which the man by hypothesis genuinely cares about. He must care about it, or he could not get the subjective satisfaction. The pursuit of subjective satisfaction in preference to objective satisfaction can lead to madness, in the literal sense of madness: you can lose your grip on *reality*. And if you think that only fanciful scenarios like the one Rawls describes could possibly give rise to that problem, think about it again

the next time you enjoy being flattered by someone whose opinion you don't actually respect.

Now at this juncture someone may wish to say instead that the rational egoist is a person who tries to maximize his objective satisfaction. But now we run into a new problem. The idea of *maximizing* objective satisfaction makes no obvious sense. Even supposing that we had some clear way of individuating and so counting our desires, nobody thinks that maximizing objective satisfaction is rational if that means maximizing the raw number of satisfied desires, for everyone thinks that our desires differ greatly in their importance and centrality to our lives. Maximizing satisfaction must have something to do with giving priority to the things that matter more to us. So we need some way of assigning *prima facie* weights of some kind to our desires or more generally to our projects before we know how to maximize their objective satisfaction. In other words, we need to know how strong a reason each of our projects provides us with, before we can combine them into some sort of maximized sum. Provided we have a theory of practical reason rich enough to assign such measures, this is certainly an intelligible procedure. But it has nothing special to do with egoism, for it is simply a procedure for determining what, given his reasons, a person has most reason to do. So I suggest we move on.

5.4.6 Next in line is the democratic person, who in contemporary jargon is kind of 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 (561b).

Democracy is a degenerate case of government, for such a person is governed 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 first. 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 to have just one, and he goes with Neil to the bar. While waiting for his beer, however, he finds that the loud noise in the bar gives him a headache, and he decides to return home without 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 look at the book, and doesn't drink his beer.

Of course the democratic life does not *have* to be like this; it is 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 is the trouble, for it is also only an accident if that does *not* happen. The democratic person has no resources for shaping his will to prevent this, and so he is at the mercy of accident. Like Jeremy, he may be almost completely *incapable of effective action*.

5.5.1 According to Plato, it is from the chaos resulting from this kind of life that the final type, the tyrannical soul, emerges. In a horrifying imitation of the unity and simplicity that characterize justice, this kind of soul is once again unified, but not by reason looking to the good of the whole. Plato tells us the tyrannical soul is governed by some nightmarish erotic desire, which subordinates the entire soul to its purposes, leaving the person an absolute slave to a single dominating obsession.

5.5.2 It's a strange moment, this bit about the tyrant. A strange entry into the ongoing argument about how we are to envision evil.

According to one view, the bad or evil person is pathetic, and powerless. The drunk in the gutter, the junkie, the stupid hothead who shoots a policeman and pays for it for the rest of his life, the perpetual loser who cannot hold down a job. Bad people are people without standards, without integrity, without plans even, who can be led in any direction by the desire or the suggestion of the moment. Bad people are people who cannot sustain friendships, because they would betray a friend for a few dollars in order to buy themselves a pleasure. Bad people are people who cannot pursue any larger or more spiritual ambitions, since their appetites always hold sway and are always diverting them from the course they set.

These cases come naturally to mind when we think of bad action as defective. For when we think of these kinds of cases we think of badness or evil as a lack, a deficiency, a psychological failure. Uncontrolled and insubstantial, the bad person cannot stay on the track of an ambition or a relationship. The good person, by contrast, is someone with standards, someone with integrity, someone who is able to govern herself. The good person is someone who can deny her appetites when it is called for by her larger purposes,

and someone who can give way gracefully to the wishes of a friend or a fellow citizen when that is the reasonable thing to do. Evil is weakness and goodness is the self-confidence of efficacious power. Call that the privative conception of evil: evil is a privation, a lack.

But then there is that other vision, isn't there? Thrasymachus's vision. According to this view, the bad or evil person is powerful, ruthless, unconstrained. The evil person is prepared to do *whatever is necessary* to get what he wants, and determined to let nothing stand in his way. He is clever enough to circumvent the law, and both able and willing to outwit, outsmart, or if necessary outshoot whoever and whatever comes between him and the satisfaction of his desire. The tyrant of the ancient Greek imagination is the glamorous Mafia kingpin of our own. So far from being *unable* to sustain relationships or projects, the evil person is more than anybody else able to stay on the track of them. For he is the one who is prepared to do *whatever is necessary*, whatever it takes. And this is where the doubt about morality comes in. Compared to the evil or ruthless person, the just and good person seems to be kind of weakling. Hedged around by rules and restrictions, the good person cannot take a single step forward without asking God or Society for permission; and the moment these forces seem to him to say no, he desists immediately. Moral rules and restrictions trap and constrain him; they impose limits on what he can do, they make him suffer agonies of guilt on the rare occasions when he does as he pleases. The good and just person is docile, tame, there to be taken advantage of by those who are stronger and more ruthless, a lamb to be led to the slaughter. Evil is power and goodness is weakness. Call that the positive conception of evil; evil is a positive force.

5.5.3 It may be said - and not I think exactly wrongly - that the work of the *Republic* is to show that the privative conception of evil is the true one. But it isn't quite that simple, for, as I said earlier, the mere privation of self-government is the democratic state. And Plato's story does not end there. The tyrannical soul, as I've just said, is consistently ruled and unified, though it is not self-governed. The tyrannical person is a slave, a terrified and captive soul, in thrall to erotic obsession; but slavery is not the mere privation of government - it's a positive state. For the modern reader, it's hard not to think of the addict, with his dominating obsession, or even more, given Plato's reference to erotic desire, of a real figure of horror from the modern landscape, the serial sex killer, condemned to the eternal reenactment of some horrifying sexual scene. So Plato evidently thinks there's something to the positive picture, something that explains its hold over us. The tyrant is not a force, but his desire is. Tyrannical desires, like tyrants themselves, see the absence of government as an opportunity, a vacancy, into which they can slip and take over.

5.5.4 But if a tyrannical desire unifies the soul, how is it different from aristocratic government? Why doesn't it make effective action possible? Well, first of all, the tyrannical person does not really choose *actions*, in the technical sense I have defined. For the tyrannical person doesn't choose *an act for the sake of an end*, the whole package as something worth doing. There's one end - as in the case of the serial killer, it may be the act itself - one end or act that he's going to pursue or to do no matter what, and it rules him. And for him that end makes anything worth doing, anything at all, and that's a fact that is settled in advance of reflection. It is this fact, the fact that he is willing to do certain things *whatever the consequences*, which makes him such an unsettling parody of the just

person. Yet our sense that there is something mechanical about him is not accidental. As I imagine the tyrant, his relation to his obsession is like a psychotic's relation to his delusion: he is prepared to organize everything else around it, even at the expense of a loss of his grip on reality, on the world. In fact tyranny is not merely like psychosis, it includes psychosis as one of its components, for as I have tried to emphasize, each principle is paired with a set of incentives, a representation of the world in its terms. The serial killer may actually see his victim as *asking for it*, for example, because he needs to see her that way; and for the addict of course the house is not full of furniture much less somebody else's furniture but of things you can sell for the money for the drug. The tyrannical person may be clever, in Aristotle's sense - he may have considerable instrumental intelligence. But he doesn't decide what is worth doing for the sake of what, he doesn't choose maxims, and that means he doesn't make laws for himself, and that means he isn't autonomous, and that means he isn't free.

5.6.1 In Plato's story, as in Kant's, bad action is action governed by a principle of choice which cannot be reason's own principle: a principle of honor, egoism, wantonness, or obsession. 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 - constitutionally unified. It is bad, because it is not reason's own principle, it does not rule for the good of the soul as a whole, and therefore the unity it produces - at least in the cases of timocracy, oligarchy, and democracy - is contingent and unstable. The agent's unity is propped, so to speak, by the fact that the circumstances that would reveal the competing factions in his soul and undercut his efficacy don't happen to occur. The timocratic person may lose track of his ends in pursuit of his honor; the egoistic person prefers an apparent

satisfaction to the very reality needed to make sense of that satisfaction; the democratic person drops his projects in the face of the slightest temptation or distraction. And tyranny, or obsession, finally, is not just a defect but in the most literal sense a perversion of self-rule, the subjection of the self to a single thing inside it.

Reason's own principle, in contrast to all of these, is the principle that truly unifies the soul, and unifies it in a way that makes it capable of effective action. And both Plato and Kant think that that principle, the one that really unifies us, and renders us autonomous, is also the principle of the morally good person. According to Plato and Kant, integrity in the metaphysical sense – the unity of agency – and in the moral sense – goodness – are one and the same property. In the next lecture I will try to explain why they think so.