Self-Constitution: Agency, Identity, and Integrity. By CHRISTINE M. KORSGAARD. (Oxford: Oxford University Press, 2009. Pp. xiv + 230. Price £45.00.)

Had Plato written in English, could he have resisted the puns inherent in our word "constitution"? In English, *a constitution* means both a political set-up and an individual's physical or psychic health, and *constitution* (the verbal noun) is the act of establishing things, including states and psyches. What could be more perfect for Plato's purposes in *The Republic*—or *The Constitution*, as it would then have been called?

There again, maybe Plato created the pun: maybe our word "constitution" now has both a political and a psychical sense *because* of the deep historical influence of Plato's city-soul analogy (and perhaps also of St Paul's church-body analogy in *I Corinthians* 12, which may itself display Plato's influence).

In any case the pun is irresistible to Christine Korsgaard. Indeed this is the second time she has used it in a book title: cp. her *The Constitution of Agency: Essays on Practical Reason and Moral Psychology* (OUP 2008), a book whose degree of closeness to the present work at times approaches fusion. (Compare, for instance, CA p.63, fn.60, with SC p.71, last paragraph, second sentence.) "I am *always* making the same argument", says Korsgaard (SC p.76), wryly combining self-parody with an enactment of her own thesis that some kinds of homogeneity or consistency stand as a necessary condition of practical identity.

Self-Constitution is nonetheless a new monograph, presenting the very latest versions of Korsgaard's central arguments. And it is a truly remarkable achievement, readable, learned, humane, and passionate. It is also beautifully written. Above all, it is *exciting*. Korsgaard is far from unscholarly, but—it seems to me—she is not afraid to be thought unscholarly because she takes risks that more costive and timid academics might eschew. Good for her.

What *Self-Constitution* takes further than any of Korsgaard's previous work is the deployment of Platonic and Aristotelian resources to Kantian ends. Chiefly, Korsgaard herself says, she uses Aristotle "to explain the sense in which an intentional movement can be attributed to an agent as its author", and Plato "to explain the kind of unity that a person must have to be regarded as the author of her movements" (SC xii). But there are other uses too. The book is as much an ingenious and novel reading of Plato's *Republic* as of Kant's ethical works.

In one line, the book argues that self-constitution is what agency is all about. The self is an achievement of reason, and action according to reason is the means whereby the self is made.

> …whenever you choose an action—whenever you take control of your own movements—you are constituting yourself as the author of that action, and so you are deciding who to be… As a rational being, as a rational agent, you are faced with the task of *making something* of yourself, and you must regard yourself as a success or a failure insofar as you succeed or fail at this task. (SC xi-xii)

> Action is self-constitution. And accordingly I am going to argue that what makes actions good or bad is how well they constitute you. (SC 25)

First question: won't action on Korsgaard's conception always be illegitimately self-regarding? The commonsense answer to "Why should I give to famine relief?" is "To help the starving". Korsgaard's answer looks to be "As part of my self-constitution". Korsgaard may answer that the two answers are always both involved—*all* reasoning is universalising reasoning; every reasoner is a Kantian reasoner, whether or not she realises it (81)—and that something goes wrong if either is not present. As other Kantians have recently insisted, there is a difference between the formal and the substantive characterisation of an action, and there need be no conflict between them. (If that seems runic, perhaps it will seem less so after a few more paragraphs.)

Second question: how, if there is only one self, can there be more than one self-constituting action? Won't the first such action leave my self created, and nothing for later actions to do? Or if a series of actions cumulatively create a self, won't the self only be complete when the series is, so that judgements about "how well they constitute you", hence about how good they are, are only possible *ex post facto*?

But what Korsgaard means by "self-constitution" is as much self-maintenance as self-creation. Our universal principles "hold us together" (103); formulating and following them can neatly be called "pulling yourself together" (125-6, 179), maybe even "getting a life" (see 128-9). Insofar as you fail to act on such principles, "you are not one person, but a series, a *mere heap*, of unrelated impulses" (76, cp. 142, 206). By acting rationally a human *keeps* himself constituted as an agent; as long as he lives, self-constitution is an "endless activity" (41). In so doing he performs what Aristotle (as Korsgaard reads him) regarded as the human function, in a sense analogous to the sense in which a giraffe performs the giraffe's function—"to be a giraffe, and to go on being a giraffe, and to produce other giraffes" (35).

Self-constitution means recognising the "reflective distance" "between the incentive and the response" (116); it means choosing to distinguish what *you* do from what impulses, instincts, and appetitions do *in* you, or *through* you. It means choosing to act not on whim but on principle. Another name for what Korsgaard means by "principle" is "reason, the thing that makes us us" (114). For unless you choose to act on principle or reason, there *is* no you: "In order to be an agent, you have to be autonomous… in order to be autonomous, it is essential that your movements be caused by you, by you operating as a unit, not by some force that is working in you or on you" (213). Thus self-constitution is not something that one action achieves, but something that *every* action should achieve.

This does clarify the issue. But as Korsgaard recognises (83), it might not see the objectors off completely. For one thing, a life where you are *always* in the business of unifying yourself sounds quite a strenuous and a high-minded life. Is there no time off, no time to be frivolous? Korsgaard's response can be Kant's: "Sure, within reason". (What boundary does reason set here? There is a familiar paradox lurking in the idea of gravely deciding to be frivolous. That paradox does not arise in practice, of course: in practice we are frivolous *without* deciding to be, simply when we don't decide not to be. But can the Kantian capture this feature of real life?)

For another thing, it is a theme familiar from, e.g., Nussbaum and Williams that it will not always be possible, let alone rational, to unify the moral life. Sometimes conflicted is the rational way to be. Korsgaard I think must deny this. She might say that the only place such conflicts can come from is a diversity of values *in the world.* Since she rejects the realism that "value in the world" implies, she will continue to align rationality and self-unification, rather as Simon Blackburn aligns them on the basis of his irrealism. I doubt this move does the trick: again like Blackburn, Korsgaard accepts that our experience of value *feels* like response to value-in-the-world, even if it isn't really (209). So simply rejecting realism will not get her round the fact that our experience of value also feels like response to *conflicting* values in the world. If the way it feels matters—and Korsgaard clearly thinks it does—the problem of conflict arises for her whether or not she is a realist.

It is tempting to put a third difficulty as follows: Mustn't there be a first occasion on which I act on principle? But how did I manage that, when I had never acted on principle *before*? How, in short, does responsibility arise out of non-responsibility?

Tempting though this way of putting it may be, it can't be right. It is entirely normal for rational competences to emerge as we develop, without our knowing exactly when and where they first appear. The question "At precisely what time, on exactly what day, were you first able to read?" need have no good answer; the ability to read appears gradually, and there is usually nothing more precise to say about its appearance than (at most) remarks like "It certainly wasn't there on June 1, and certainly was there on July 1". If we can allow this vagueness in the acquisition of other rational capacities, we can allow it in the acquisition of reason or principle.

Korsgaard does have a developmental story to tell about how we acquire reason, but (to her credit) she does not lose sight of this inherent and constitutive vagueness. The title of her sixth chapter, "Expulsion from the Garden", suggests that the form of this story will be frankly mythical. However, its real point—for that matter, perhaps the same can be said about the Garden of Eden story—is about how humans differ from animals (212-3):

> A non-human animal acts on what I called "instinct". Her instincts are her principles, and they constitute her will… by structuring her perceptions [so that] she nearly always already knows what to do… You are not so lucky. As a rational agent, you are aware of the grounds of your beliefs and actions—or, I should say, the potential grounds. For being aware of them gives you some distance from them… Self-consciousness divides you into two parts, or three, or [more]… it separates your perceptions from their automatic normative force… On the one side, there is the [inclination to run;] on the other side, there is the part of you that will make the decision whether to run, and we call that reason. Now you are divided into parts, and must pull yourself together by making a choice. And in order to make that choice, reason needs a principle—not one imposed on it from outside, for it has no reason to accept such a principle, but one that is its own.

Self-consciousness enables us humans, unlike the giraffe, to be *aware* that we have an impulse to run. To be aware that we have this impulse is to be able to be "reflectively distant" from it. Reflective distance from the impulse to run makes possible the deliberative question "But *should* I run?". That question, once raised, demands an answer; it brings with it "the knowledge of good

and evil" (Genesis 3.22). But the deliberative question can only be answered by reference to a "principle": that is to say, by reference to a rule for acting in *any* such case.

Such principles could be imposed on us from outside (heteronomously, as Kantians say), but then they would not be *our* principles. The only remaining possibility is that we should create such principles, "autonomously", for ourselves, to constitute the form of our deliberation on any given occasion. But, if I am to overcome the fragmentation created by self-consciousness, the principles that I create for myself in this way must be *self-unifying* principles, principles on which it can be truly said that *I* act, not merely that *some force in me* acts through me. And only some principles can be self-unifying principles. For example, a principle always to act on whatever whim happens to seize me at the moment—"particularistic willing", as Korsgaard calls it (72-6)—leads not to psychic unity, but to disintegration. (*The Constitution of Agency* (59, n.52) gave us the nice example of Jeremy the aimless student, who is always halfway through one doing thing when another stimulus comes along and he starts doing something else; Korsgaard reuses this at 169.) Other 'principles' are not really even principles. For example—in this book, Korsgaard's favourite example—it is not just self-stultifying, but actually incoherent, to try to determine whether you should tell yourself the truth on pragmatic-utilitarian grounds (182-3):

> [If] you undertake to believe, not what is true… but what it is most useful for you to believe[, then] you've got a problem… before you can allow yourself to believe what is true, you have to satisfy yourself that it is most useful for you to believe what is true. But you can't do that without first satisfying yourself that what you think about *that*—about what it is most useful to believe, I mean—is, quite simply, true… So you have to try to tell yourself the truth… You can't treat yourself in accordance with the principle of utility while you are thinking. For the principle of utility is a tyranny, while, thought, by its very nature, is free.

Our agency is determined by our principles, and our principles are the principles we choose. But choice of principles must be choice of *self-unifying* principles. And some would-be principles, such as Jeremy's, are self-disintegrating not self-unifying; while others, such as the utilitarian's about truth-telling, fail to be principles at all.

In the main contours of this argument, Korsgaard takes herself to be expounding both Kant and Plato. As she rightly says, the key theme of the *Republic* is the reunification, into an integrated and harmonious whole, of the parts of the soul: parts that only become separate in the "expulsion from the garden", in the moment when self-consciousness makes us see our own desires and impulses as something we can *stand back from*. (The first quotation in the book is Plato's marvellous words from *Republic* 443d about the just man's self-harmonising.) She must be right too to see this notion of reflective distance as key to Kant's distinction between reason and inclination, and to claim that reason, for Kant, is crucially an organising and harmonising faculty. Hence what Kant has to say about *reason* can be very naturally and fruitfully combined, as she combines it, with what Plato's *Republic* has to say about *justice.*

Another sign that her combination of the *Republic* and Kant is an apt one is this: on her reading, the same basic difficulty faces both. This is the intuitive gap between self-constituting action and morally good action. If Gauguin walks out on family life to go to Tahiti to paint, he follows the imperative to constitute himself according to his own preferred principles, but (apparently) acts

immorally; if Gauguin stays with his family, he does the moral thing, but (apparently) misses the best way to advance his own self-constitution. For Kant, this is the familiar problem whether the principles that the categorical imperative generates—if any—are the same as the principles of morality. For Plato, it is the equally familiar problem whether "philosophical justice" is the same thing as "vulgar justice". Korsgaard's achievement is to shed some light—*some* light—on both puzzles by taking them together.

I say "some light", because in some cases it seems very clear how Korsgaard might bridge this intuitive gap; in other cases, not clear at all. For example, the argument quoted above for the incoherence of pragmatism about truth-telling *towards yourself* is only half of the case made at 182-3. The other half is about the incoherence of such pragmatism *towards others*, and the key thought that ties the two halves together is "there is really no reason for you to treat yourself any differently than anyone else" in respect of truth-telling (182). So if the pragmatic maxim about truth is incoherent as applied to me, it must be incoherent as applied to others too.

So the argument will go. I am not endorsing it, nor rejecting it; I am merely saying that I can see what it will be. I find this harder to see in the Gauguin case. As above, Gauguin's choice of Tahiti over family life apparently increases his *Platonic* integrity (as we might call it), but decreases his *moral* integrity. That looks like a proof that Platonic integrity and moral integrity, philosophical justice and vulgar justice, are two different things—even if, as perhaps in the case of truth-telling, they sometimes coincide.

The Gauguin case is particularly tricky, because in it moral and Platonic integrity, as I have called them, actually pull against each other. In Korsgaard's defence, it might be pointed out that in most cases they do not conflict. Jeremy the aimless student illustrates nicely how a lack of principle can lead to ineffectuality. You might think that the weakness of Korsgaard's argument here is that wickedness has nothing to do with ineffectuality: that the wicked are typically as frighteningly dynamic, as deadly in the effectiveness of their agency, as "good at being persons" (xiii)—only in the wrong way—as Plato's Thrasymachus, or Milton's Satan, or Shakespeare's Richard III, or *The Godfather's* Vito Corleone.

But these are all *fictional* characters. Turn to real life, and the link between wickedness and ineffectuality is far clearer, in even the most supposedly Satanic cases. Someone might rhetorically ask, "Was Hitler ineffectual?", expecting the answer "Of course not" and the collapse of Korsgaard's case. But actually the answer is "Yes". Hitler didn't get up till 10, and spent his mornings ranting pointlessly at his generals, and his afternoons asleep.[1] Nor was Hitler's deputy Himmler much more effective: he had no idea how to run a government or an army, and spent as much as he could of his time designing uniforms and medals. In real life, highly effective people who are also deeply wicked are actually remarkably rare. The tragedies of history, it seems, happen mostly not when effectiveness and wickedness are combined in the same people. Rather they happen when, for whatever reason, the efficient and honest people do what the wicked and ineffectual ones tell them to.

[1] This was certainly how he behaved in his later years. Perhaps he was more effectual in the 1930s—and perhaps he was less evil then. (Thanks to Derek Parfit for querying this claim.)

But the Gauguin case, to come back to that, is not about wickedness as such; it is about how there might be a difference between Platonic and moral integrity. Evidence about what Korsgaard might say about the Gauguin case comes principally, within *Self-Constitution*, from her discussion of Parfit's Russian nobleman (185-6; cp. Parfit, *Reasons and Persons* 327-8). The nobleman is a socialist in youth, a "reactionary" (as Russians like to call conservatives) in old age. In his relatively penurious youth he foresees that "he will have different values" when he is old and rich. Anticipating this, he now draws up a plan for land emancipation in the future. The plan cannot be revoked without his wife's permission, and he binds her never to give this: he tells "his wife that his younger self is his real self, that his ideals are essential to him, and that if he loses those ideals she should regard him as… dead [and so] unable to release her from her promise" (185).

What does the Russian nobleman have to teach us about self-constitution, hence about moral and rational action? That wherever it is possible to *delegate* responsibility for your future behaviour it is also, and *a fortiori*, possible to *take* responsibility for it. The wife "is to hold him, by holding herself, to giving up the estates. But if she can do this, why can't he?" (187). Unless the nobleman's agency is radically impaired, e.g. by mental illness, this is a rhetorical question. He can. His foreseen future attitudes are not merely matters of *prediction* for him, as someone else's attitudes might be. They are one of the things that he, now, has to bring within the scope of his deliberation now. Instead of just anticipating his older self's "reactionary" attitudes, he has to *look into the content* of those attitudes: to decide whether they are justified attitudes, and so whether it is his future conservative self, or his present socialist self, who is right, and hence the self that he really identifies with, not in some imagined future, but right now in the actual moment of his deliberation (204). A more perfect illustration of what Korsgaard means by her claim that deliberation is necessarily self-unifying, both synchronically and diachronically, would be hard to imagine.

A parallel line of reasoning would presumably apply to Gauguin. What stands in the way of his deserting his family and going to Tahiti to paint? Though Korsgaard has interesting things to say about marriage (186-191), her answer is not necessarily—as the *Groundwork*'s surely would be—just to point out that he promised, on the day of his marriage, to stay with his wife and any children there might be. Promises can be made rashly, or on crucially incomplete information, or in the grip of youthful (or other) illusions about oneself or others or about "Life"; Korsgaard is not a fetishist about promises in the way that some Kant-inspired moralists, and indeed Kant himself, sometimes are. Possibly *nothing* stands in the way of Gauguin's going to Tahiti. It depends which Gauguin Gauguin will decide is the real one, if he gets the decision right: the Gauguin who accepts that past marital promise as the expression of his true self, or the Gauguin who rejects it as expressing some misunderstanding of what he is trying to "make of himself" (xii).

This is what I think Korsgaard might say about Gauguin. I can only apologise if I've misunderstood her. Misunderstood or not, her argument suggests an extremely appealing and interesting way of dealing with such cases. More widely, the general importance in ethics of her great theme of self-unification should be clear. Any academic who arrives at his desk in the morning to a babel of emails and to-do items pointing him in forty different directions at once is

bound to see the crucial link between self-integration and effectiveness in doing what you actually *want* to do: in Kierkegaard's famous words, "purity of heart is to will one thing".

My main doubt about such an argument is simply whether it is a *Kantian* argument. An argument about self-expression, self-development, and authenticity, and about the kind of consistency that these ideals require, seems at least as much an existentialist argument as a Kantian one. That is not necessarily a criticism; perhaps an existentialist argument is what we need. But it is a signal that the Kant we are now dealing with, Korsgaard's Kant, is a very different character from the stern, pietistic, moralistic rigorist of the tradition. It is also a signal that the original question, whether the ideal of self-constitution lines up with the precepts of morality, remains unanswered. One lesson we might draw from Korsgaard's wonderful book is that in fact the two ideals do *not* line up—and that is just so much the worse for the precepts of morality.

Many other fascinating questions, large and small, are raised by Korsgaard's wonderful book that I cannot discuss here. Among the small questions, I am keen to make time soon to think further about what a bad house-builder aims at, if not a house (31-2); also about whether I am alone if there is just me and a goldfish in the room, as I certainly am not alone if it's me and a dog, and certainly am alone if it's me and a daffodil (129). Among the large questions, one to return to is the question how far Korsgaard really is from the realism that she so often impugns. She talks in her irrealist way as if it would be heteronomy, a kind of idolatry, to live on realist assumptions. Yet her own view, simply put, is both that we create values (25, 209) by choosing what to value ("making the contingent necessary", 23); and also that there is basically only one right way to make these choices (71-2). Why is that so, if not because the power to self-determine is subject to rational constraints which themselves are written into the nature of things? But if there are rational constraints written into the nature of things, then realism of some sort is true. If one of the things that those rational constraints are written into is the human heart, that is not a refutation of realism but an application of it, which will show how autonomy and realism are consistent after all. For then moral commands (like God's commands on the best solution of the *Euthyphro* dilemma) will come not only from within, and not only from outside, but from both places: from nowhere and from everywhere.[2]

[2] For comments and encouragement, thanks to Michael Brady, Derek Parfit, and Christine Korsgaard herself.