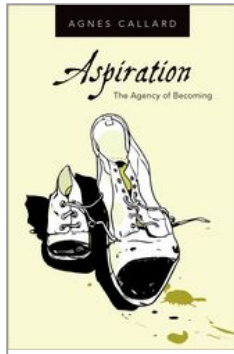


University Press Scholarship Online  
Oxford Scholarship Online



## Aspiration: The Agency of Becoming

Agnes Callard

Print publication date: 2018

Print ISBN-13: 9780190639488

Published to Oxford Scholarship Online: March 2018

DOI: 10.1093/oso/9780190639488.001.0001

## The Problem of Self-Creation

Agnes Callard

DOI:10.1093/oso/9780190639488.003.0006

### Abstract and Keywords

The new values, acquisition of which constitutes my act of self-creation, must be either continuous or discontinuous with the ones I already have. If they are continuous, I am not changing but rather working out the implications of the person I already was. If they are discontinuous and the new values contradict or come at a tangent to my old values, the change is not a product of my agency. I change, but I do not change myself. This paradox, adapted from the work of Galen Strawson, can be solved if we allow that the direction of value-dependence may be teleological: the aspirant's values depend on, and are entailed by, those of the person she is trying to be. The aspirant does not fashion, control, or make the self she creates. Instead, she looks up to that self, tries to understand her, endeavors to find a way to her.

*Keywords:* Galen Strawson, self-creation, Christine Korsgaard, teleology, source of normativity, G. E. M. Anscombe, self-cultivation, self-endorsement

In part I (chapters 1 and 2), we considered aspiration from the vantage point of decision theory: How could it be rational for someone to pursue a project that is not the optimal way of satisfying her current set of preferences? Our answer was that the rationality of aspiration cannot be captured within the synchronic framework of decision. The aspirant's reason is proleptic, with a proximate face that speaks to the person she is now and a distal face fully visible only to the person she will become. Proleptic rationality is essentially extended in time, being a form of value-learning.

We then turned in part II (chapters 3 and 4) to the moral psychology of aspiration, attempting to give an account of what proleptic rationality feels like. The aspirant is trying to see the world through another person's eyes, namely, through the eyes of the person who has the value she aspires to acquire. Her condition cannot be captured in the Frankfurtian framework: she neither "identifies with" nor is "alienated from" the evaluative perspective she imperfectly inhabits. Rather, she is in a state of intrinsic conflict between the evaluative perspective she seeks to acquire and the one she seeks to depart from.

In this final part of the book (chapters 5 and 6), we will examine the implications of the phenomenon of aspiration for the theory of **(p.180)** autonomy. The aspirant does not *end up with* a new value in the way that one might end up with an ulcer or an inheritance. She orchestrates her value-acquisition, driving herself toward a different value-condition from the one she is in. From this vantage point, aspiration emerges as a kind of *work*. The work is visible in her struggles to sustain interest in the hobby or relationship or career or religion or aesthetic experience that will later become second nature; in her repeated attempts to "get it right," attempts that must be performed without the benefit of knowing exactly what rightness consists in; and, most generally, in the fact that she always wants and strives to be farther along than she is. An aspirant's value-transition is her own work, which means that she is a certain kind of cause: a cause of herself.

This is not to deny that the aspirant receives help. Major value-acquisitions reflect the influence of one's environment, especially the people in it. Parents, teachers, and lovers have transformative effects on the people they parent, teach, and love. But they cannot, at least typically, have these effects on someone unless she participates in the process. They assist, rather than substitute for, the activity of the agent herself. My aim in these two chapters is to offer an account of the work an agent does in making herself into a person with new values. I call this work "self-creation."

That label may sound overblown, given that the person must have been around to do the creating. Nonetheless, he didn't then exist *as* the person he becomes. The advantage of "self-creation" over "self-change" or "self-revision," as I might also have called it, is that the former term serves as a reminder that we are not interested in minor or superficial ways in which someone might change himself. Self-creation in the sense I'm interested in is not going to be a matter of a person making physical changes to his body, altering his musculature or the color of his hair. Caring about something **(p.181)** new can have physical repercussions, but the change one is effecting on oneself is not, in the first instance, a physical change. Nor is it a psychological one. It will, to be sure, have psychological repercussions: one's new form of valuation lends itself to new emotions, feelings of attachment and vulnerability, curiosity, interest, excitement. But these, too, are consequences of something deeper, which is the *ethical* change occasioned by committing oneself to some form of valuation. Our interest is in self-directed value-acquisition, which is, first and foremost, a change of a person in the ethical dimension. In this chapter, I propose to explain how the aspirant can be an ethical source of himself.

I begin with a discussion of why this question is so difficult to answer. I describe two received models for self-directed agency and argue that neither offers us a framework for understanding self-creation. First, I consider the possibility that one creates oneself by an act of the will in which one endorses some way of being, sanctioning it with an evaluative stamp of approval. I call this the “self-endorsement model,” and I argue that because one is in a position to endorse only what one *already* values, self-endorsement cannot represent a way of *acquiring* values. Next, I turn to what I call the “self-cultivation model.” Self-cultivation is the process of working to satisfy anterior normative commitments about what kind of person to become. I adapt a paradox described by Galen Strawson (1994) to show that we cannot tell a story of genuine self-creation as a story of self-cultivation. The paradox runs as follows. If, on the one hand, the value I cultivate in myself follows rationally from values I already have, then I do not do any creating. For in this case my “new” self was already contained in my old self. If, on the other hand, the new value is rationally unconnected to my earlier values, then its advent in my life cannot be my own doing. In this case, the self I end up with may be new, but it is not the product of my own **(p.182)** agency. Rational value-cultivation is not *self-creation*, and nonrational value-cultivation is not *self-creation*.

I analyze the underlying problem as follows. Call the self that does the creating “ $S_1$ ,” and call the self that is created “ $S_2$ .” For the theorist of self-endorsement,  $S_1$  is the endorsing self,  $S_2$  the self it endorses; for the theorist of self-cultivation,  $S_1$  is the cultivating self,  $S_2$  is the self it cultivates. Both theorists depict  $S_1$  as normatively prior to  $S_2$  in the following sense: they present  $S_1$  as the agent’s authoritative self, the self whose agency (via endorsement or commitment) determines how  $S_2$  ought to be. I argue that if we attempt to derive the norms governing  $S_2$  from the values or commitments already present in  $S_1$ , we will never be able to describe self-creation. Instead, one must reverse the priority relation between the two selves. In section IV of this chapter, I show that we can do so by conceiving of the temporally posterior created self as authoritative over the self who creates her. The aspirant does not see herself as fashioning, controlling, sanctioning, making, or shaping the self she creates. Instead, she looks up to that self, tries to understand her, endeavors to find a way to her.

Skeptics of self-creation, Strawson among them, are wont to cite with approval a passage in which Nietzsche scoffs at the philosophical impulse “to pull oneself up into existence by the hair, out of the swamps of nothingness.” He describes the very idea of self-creation as a “rape and perversion of logic.”<sup>1</sup> Such skeptics may worry that an account such as mine succumbs to the basic error Nietzsche is describing: in positing the normatively prior self as temporally posterior, I am presenting a teleological account of (a species of) **(p.183)** self-directed agency. Is such teleology naturalistically suspect—does it rest on some notion of backward causation or causeless effects?

I argue that the primary task of someone responding to Strawson’s dilemma is that of giving an account of how one set of norms—the norms governing  $S_1$ —is related to another set of norms—the norms governing  $S_2$ . This question about normative grounding needn’t be identified with the question as to the *causal* grounding relations between  $S_1$  and  $S_2$ . Nothing prevents the theorist of aspiration from offering a traditional (i.e., non-teleological) account of the causal grounding of the genesis of  $S_2$  in the representations, desires, etc. of  $S_1$ . But if we want to know whether those very desires and representations succeed or fail, we must assess them with reference to the as yet nonexistent  $S_2$ .

In aspiration, it is the created self who, through the creator’s imperfect but gradually improving understanding of her, makes intelligible the path the person’s life takes. Aspiration is that form of agency in which one acts upon oneself to create a self with substantively new values. One does this by allowing oneself to be guided by the very self one is bringing into being.

### I. Self-Endorsement

A number of philosophers have taken an interest in our power to endorse or withdraw approval from some feature of ourselves. Harry Frankfurt distinguishes between two kinds of drug addict: the unwilling addict rejects his own impulse to take a drug, by contrast with the willing addict who approves of his own addiction. Frankfurt (1971) says that the former does not desire as he *desires to desire* and (in Frankfurt 1988) that the latter *identifies* with **(p.184)** his desire. Christine Korsgaard (1996) might describe that same addict as endorsing (or failing to endorse) the practical identity of being an addict (ch. 3). Gary Watson covers the same territory by contrasting a person's evaluative and motivational systems: the former issues judgments as to the value of the desires populating the latter. If approval is withdrawn, as in the case of the unwilling addict, the agent is "estranged from his . . . inclinations" (1975: 210).

Abstracting from their differences, we can group these views together as describing *self-endorsement*, i.e. an activity in which an agent, or some part of her, steps back from, appraises, and attaches a positive or negative evaluation to the aspect of herself she evaluates. I do not doubt that self-endorsement is a real and significant aspect of agency, but I will argue that it cannot serve as a means of self-creation. We do not change who we are by reflectively approving or disapproving of ourselves.

Self-creation, as I understand it, involves the creation of values. But values are also what we use in forming our endorsements or rejections of some feature of ourselves. Can one use one value to generate or eliminate another? Consider some possibilities for how this might go. Suppose that I have exactly three values:  $V_1$ ,  $V_2$ , and  $V_3$ . I cannot simply ask myself, “Is  $V_1$  valuable?” for then I will be putting my finger on the scales. I *do* value  $V_1$  and will end up evaluating  $V_1$  from a point of view that assumes the value of  $V_1$ . But I might successfully set that value aside by asking myself a hypothetical question: “Should someone who values only  $V_2$  also value  $V_1$ ?” or “Should someone who values only  $V_2$  and  $V_3$  also value  $V_1$ ?” In this way, I can investigate the question of whether to value  $V_1$  without presuming its value. But notice that if I answer no to both of these questions, this has no implications for what *I* should do or value: I am not someone who values only  $V_2$  or who values only  $V_2$  and  $V_3$ . The hypothetical model allows an agent to evaluate one value from **(p.185)** the point of view of another, but at the expense of being uninformative of her actual situation. Such hypothetical reasoning does not put the agent in a position to change her value-endorsement.

This observation might lead the theorist of self-endorsement to abandon the project of evaluating a value by way of some value set that does not include it, in favor of evaluating a value by looking at its place in one’s value-system as a whole.<sup>2</sup> One can observe, while still valuing  $X$ , that one’s valuation of  $X$  does not cohere well with the rest of what one values.

Suppose, then, that someone realizes that one of her values—say, her valuation of her appearance—detracts from her ability to value something else—say, equality or independence—to which she is more fundamentally committed; or that given her feminist values she really ought to value a certain form of political activism to which she is currently indifferent. Do we, in this case, have a form of self-endorsement that amounts to self-creation?<sup>3</sup>

No: the realization in question stops just short of the relevant change. We are to imagine the agent asking herself a question **(p.186)** such as “I know I value *X*, but *ought* I to value it, given my valuation of *Y* and *Z*?” or “I know I don’t value *W*, but *ought* I to value it, given my valuation of *X*, *Y*, and *Z*?” If she concludes that she ought to value *W*, this does not constitute an adoption of that value, but rather a change in her judgment as to what she ought to value. Such an agent comes to see that she is not as she ought to be. The project of assessing oneself for coherence presupposes the possibility of separating the question of what values one has from the question of what values one should have. That separation characterizes the agent who answers the question, for she concludes that she shouldn’t (or should) have a certain value *that she continues to have* (or not have).

The judgment that she would be more coherent if she valued *X* cannot itself constitute a change in what she values. Nor can we say that the formation of the judgment that I ought to value *X* amounts to the creation of a second-order value. If there is such a thing as valuing a value, this value cannot be what is produced by my coherentist self-policing. Rather it must, like my first-order valuing of *X*, figure among the objects I police. When I ask myself what I must value in order for my values to be coherent, I step back from my higher-order values as well as my lower-order ones, since I am asking myself not what I *do* value, or what I *do* value valuing, but what I ought to value, or ought to value valuing.<sup>4</sup>

**(p.187)** It might seem, however, that coherentist self-policing is but one small step away from value-change. The realization that it would be rational to gain or lose a value may not itself amount to value-change, but it could constitute an impetus to value-change. In the face of her acknowledgment that she is not the person she ought to be, the agent might be moved to go ahead and change (or commit to changing) herself into that person. In the next section, I will consider whether we can model self-creation using the process of bringing oneself into line with one’s judgments about one’s values. This process includes, but extends beyond, the moment of self-assessment to encompass the agency that flows from and is governed by whatever assessments one makes.

## II. Self-Cultivation



In chapter 1, I described a form of preference change that I called “self-cultivation.” The person who cultivates herself engenders preferences in herself whose value she antecedently appreciates. For instance, her interest in living a healthier life moves her to join a gym, in the expectation that she will thereby, over time, engender in herself an inclination to exercise. When we cultivate ourselves, we don’t just stand back and assess ourselves; we actually try to change ourselves. Thus self-cultivation might seem to be a more promising model for self-creation than mere self-endorsement. Philosophers such as Jean Hampton, Michael Smith, Michael Bratman, Richard Holton, and Joseph Raz have all articulated what can, broadly speaking, be characterized as theories of self-cultivation. For they all describe a process by which a person can act so as to determine her future self.

**(p.188)** Hampton describes a “self-authorship” process: “There are many times in our lives when we choose what we will be. For example, when a young girl has the choice of entering into a harsh regimen of training to become an accomplished figure skater, or else refusing it and enjoying a more normal life with lots of time for play, she is being asked to choose or author whom she will be. When a graduate student decides which field of her discipline she will pursue, or when a person makes a decision about his future religious life, or when someone takes up a hobby—all of these choices are ways of determining one’s traits, activities, and skills, and thus ways of shaping one’s self—of determining one’s self-identity” (1993: 150). Hampton contrasts this “authentic” form of self-determination with one in which a person allows herself to be shaped by social or environmental pressures into becoming what people want her to be. “Self-authorship involves . . . a decision to develop the traits, interests, and projects that are not only consistent with meeting your objective human needs but that are also ones you want, and not ones that others prefer that you want (and perhaps try to persuade you to want)” (155). Thus she grounds the process of self-authorship in the desires of the authoring or creating self.

Smith imagines self-cultivation as a reflective process, akin to the coherentist self-policing described earlier, by which one moves oneself toward a condition of greater consistency among one's desires: "Systematic reasoning creates new underived desires and destroys old. Since each such change seems rationally required, the new desiderative profile will seem not just different from the old, but better; more rational. Indeed, it will seem better and more rational in exactly the same way, and for the same reasons, that our new corresponding evaluative beliefs will seem better and more rational than our old ones" (1995: 116). Bratman (2003) develops Frankfurt's higher-order model in a diachronic direction. He (p.189) presents the adoption of higher-order self-governing policies as a solution to the problem he calls "underdetermination by value judgment," i.e., the situation in which one is confronted with a choice among what are, from the point of view of one's antecedent desires, a multiplicity of equally valuable pursuits. When there are many good things you could do, committing yourself to one can settle which you subsequently ought to do. Holton defends the possibility of "rational non-reconsideration" (2004: 3), by which a person adheres to earlier resolutions as to what she should now do and does not open the matter for deliberation. Raz (1988), whose view I will describe in more detail later in the chapter, describes the agent at the output of the process of cultivation as doing something akin to keeping promises made by her past self.

In all of these cases, an agent makes and subsequently lives under the normative guidance of some kind of choice as to how she should live.

We certainly do cultivate our physical appearance, friendships, desires, hobbies, traits, skills, traits, activities, and careers. But the difficult case is the cultivation of the state of valuation from which the agent will select *this* physical appearance, *that* friend, rejecting *this* desire in favor of *that* one. Can someone cultivate value in herself? In the course of an attack on the possibility of moral responsibility, Galen Strawson (1994) provides the materials for an argument that value cultivation is subject to the following dilemma.<sup>5</sup>

**(p.190)** The new values, acquisition of which constitutes my act of self-creation, must be either continuous or discontinuous with the ones I already have. If they are continuous, my new values are entailed by my old ones. In this case, I don't really change. If they are discontinuous, the new values contradict or come at a tangent to my old values. In this case, they arise accidentally or through external influence rather than through my agency. I change, but I do not change myself.

Let us see how this dilemma plays out in a schematic case. Suppose that I value *E*, and I come to value *M* because I see that *M* is the means to *E* (or, more generally, that valuing *E* rationally requires me to value *M*). Will this count as self-creation? No: instrumental reasoning works out the consequences of the value condition I already have. It does not discover ends, but only means. But what if, prior to realizing the relationship between *E* and *M*, I actively disvalued *M*? Overcoming my hatred of *M* might then constitute a substantive change—a change in my ends. But in that case, the mere discovery that valuing *E* rationally required me to value *M* is insufficient to explain the change. For why didn't I abandon *E* instead of embracing *M*? Perhaps I have *other* values that dictate this decision, or perhaps I simply valued *E* more than I disvalued *M*. In either case, we are once again back on the first horn: this isn't a change in value but merely a working out of the consequences of values I already have. If I neither valued *E* more than disvaluing *M* nor had other values dictating the overall importance of *E*, then we have not yet explained my rationale for adopting *M*. I may have done so randomly or on a whim, but if my new self is the product of such arbitrariness, it is not something I made. Strawson makes this observation when considering a libertarian defense of free will. He asks, "How can the occurrence of partly random or indeterministic events contribute in any way to one's being truly morally responsible either for **(p.191)** one's actions or for one's character?" (1994: 18). This point applies equally to cases in which one's efforts of self-change are made for no reason, or arbitrarily. As Strawson points out, someone who succeeds as a result of such an event is "merely lucky." He has ended up with a good self that he did not create.

At each point in the story of  $E$  and  $M$ , we face the same choice: if the value I engender in myself follows rationally from values I already have, then I do not do any creating. For in this case my “new” self was already contained in my old self. But if the new feature is rationally unconnected to my earlier values, its advent in my life cannot be my own doing. In this case, the self I end up with may be new, but it is not the product of my own agency.

Why does the fact that the  $t_2$  value was derivable from the agent’s  $t_1$  values disqualify the case from counting as self-creation? We should allow that the derivation might have required time and effort, so that the agent at  $t_2$  differs substantially from the agent at  $t_1$ . Coming to value something that is instrumentally conducive to ends one had all along might occasion observably significant changes in acting, thinking, and feeling. But if we allow that such an agent develops her new self by unraveling the implications of the materials present in her old one, we will find that we have simply pushed the question about self-creation back to an earlier stage: How did she acquire those materials? And so the problem now is not that we must deny that the agent at  $t_2$  is self-created; the problem is that the process seems to have gotten going before  $t_1$ . Her valuation of the means is a *new expression* of a value that she acquired at whatever time (before  $t_1$ ) she came to value that end. We will not have an account of how her ethical self came into being until we have a story of *that* genesis. But if we try to tell one, Strawson’s dilemma recurs.

We encounter the same regress if we press, on Bratman's behalf, the question of why a value-choice underdetermined by one's **(p.192)** previous commitments cannot constitute self-creation. Suppose Sartre's resistance fighter was not antecedently more committed either to his mother or to his country. Forced to choose, he decides (at  $t_1$ ) to devote his life to mother over country. Does it really follow from the fact that it would have been equally rational (at  $t_1$ ) to opt to fight that the filially pious self emerging at  $t_2$  is arbitrary in relation to his self at  $t_1$ ? Bratman might point out that this would follow only if at  $t_1$  he cared little or nothing for his mother. Suppose, to the contrary, that the soldier at  $t_1$  already cared deeply for both mother and country. In this case the value must already have been created sometime before  $t_1$ . The decision made at  $t_1$  is then not value-creation, but one (of at least two) possible developments of his antecedent state of motivation. If we want to know how *that* state got created, we will have to look back further into his past. And then Strawson re-poses his dilemma.

Whenever we tie some new value to an act of self-cultivation, we must admit the arbitrariness of the choice or push the source of the cultivation back a further step. And this regress is a vicious one: as we retreat backwards through a person's selves, we encounter selves that are less and less, and eventually not at all, in a position to do any creating. And shortly thereafter, of course, we run out of self altogether.

Must my later self either come at a tangent to my earlier self or be a rational consequence of it? If the two horns of Strawson's dilemma were exhaustive of the possibilities, self-creation would be—as Strawson thinks it is—a chimera. I will argue that they are not, but I want to note that Strawson's conception of the space of possibilities is widely, though not always explicitly, shared by theorists of the self. Consider, for instance, Robert Noggle's description of "the basic picture of the self" that is "quite common among philosophers" (2005: 108 n. 26). Such a self "evolves according to its **(p.193)** own internal logic—its own contents determine whether and how it is to change in response to new information, internal conflicts and changing condition . . . when psychological changes happen in this way, it seems correct to say that the new configuration is an authentic continuation of the previous configuration. On the other hand, a psychological change—especially a change in the core attitudes—that does not occur in this way produces a new configuration that is not an authentic continuation of the previous one . . . If the changes are radical enough, it might be proper to speak of the destruction of one self and its replacement by a new one" (100–101).

As examples of changes in the second category, Noggle offers "sudden organic trauma" and "nefarious brain surgery." He contrasts these with changes that are "internally motivated in such a way that they seem to be *intelligible reflections* of the contents of the core attitudes (100, emphasis in original)." Noggle's division recapitulates Strawson's dilemma: he sees changes in a self as either rational extrapolations of previous conditions ("according to its own inner logic") or traumatic intrusions from without. Noggle's discussion of self-creation showcases how deeply we have been gripped by Strawson's dilemma. The two selves must be rationally connected, or we cannot see the second as the work of the first; but this same continuity prevents novel values from arising.

### III. Neurath's Boat

At this point, we may be tempted to conclude that self-creation is, indeed, impossible. Why should we demand that we be able to do more than revise our selves in the light of our current values, becoming more consistent versions of the people we are? We may be inclined to heed Nietzsche's advice to stop trying to pull ourselves **(p.194)** "up into existence by the hair, out of the swamps of nothingness." In order to pull ourselves up by the hair we would have to be able to detach ourselves from our bodies. And this may seem to be precisely the logical problem at the heart of the idea of self-creation: we cannot step outside ourselves. It has become a commonplace to describe the developing self using the image of Neurath's boat, which must be rebuilt as it sails.<sup>6</sup> It is impossible to fashion oneself as though one were not already some person with desires and judgments— moreover, with desires and judgments that speak to the question of how one ought to fashion oneself. One can, at most, refashion from within. I want to take a moment to motivate the claim, which may at this point seem questionable, that there is indeed a *phenomenon* of self-creation for which we should try to account. Let me begin by pointing out that Neurath's boat is not Theseus's ship.

The fact that Neurath's boat never docks is indeed an elegant way of expressing the unsheddability of character. One cannot "operate" on one's self from the outside, as a doctor might operate on his own anaesthetized leg. One cannot set aside one's way of looking at the world and somehow examine things from the outside of one's self; as we saw in our discussion of the power of reflection in chapter 3, reflection does not afford a person the opportunity to "step back" from her values. But it does not follow from the fact that reflection has limited power that the self must stay true to its initial form or else be traumatically violated from without. Neurath's boat isn't Theseus's ship, whose job is to stay the same through all its many changes. On Theseus's ship, a rotting plank must be replaced by a plank of a similar shape and size. On Neurath's boat, the only requirement is that it be replaced by sailors on board the boat. Neurath's rowboat might be gradually rebuilt into a trireme **(p.195)** or, for that matter, an airplane without the builder ever setting foot onto dry land. The fact that there is no vantage point one can simply adopt outside one's character doesn't entail that one couldn't *arrive* at the vantage point that is outside one's current character by working toward that condition.



My claim, in effect, is that we are not stuck repairing rowboats; that we regularly build triremes and airplanes out of ourselves. We do this by becoming spectators, taking classes, doing exercises. We find mentors to emulate or fellow travelers with whom to commiserate—and compete. We do the same thing over and over and over again until we get it right, without knowing in advance what “right” is. We do work we don’t always enjoy, and we pretend—even to ourselves—that we enjoy it. We leave ourselves open to certain kinds of experiences and closed to others, knowingly risking disappointment and disillusionment down the line. We alert ourselves to and steel ourselves against temptations to abandon course in favor of a more readily available and more immediately intelligible form of value. Candy, television, alcohol, a nap, video games, internet surfing—pick your poison; it’s waiting in the wings. We struggle against implicit or explicit messages, from individuals or groups of individuals, to the effect that this kind of value is “not for you.” Often these struggles are heightened by the fact that we have internalized the judgments in question. The work that we are engaged in is the work of bringing something into view. But because what we are bringing into view is something practical—a value—the work is a matter of acting and feeling, as well as thinking.

This kind of work involves both moving toward and moving away from a perspective on value. When engaged in it, not only are we gaining something, we are also often losing something. In some cases, the value-perspective that we are losing can be characterized relatively thinly as the non-appreciation of something. But **(p.196)** sometimes agents can identify strongly with their pre-aspirational valuation condition. Sometimes people view their own indifference to art or music as connected with being unpretentious and straightforward. Adults who describe themselves as “child-free by choice” view their condition not in terms of the absence of children, but in terms of the presence of the freedom to dedicate themselves to work or friendship or romance or travel. The bitter wife is not merely lacking in love; she actively *hates* her husband. Perhaps she feels, given the way he treats her, that it is only by withholding her affection from him that she can respect herself. Alcibiades sees Socratism as the loss of the only life he knows how to lead, the life of honor. People in these conditions have a strong barrier to changing in respect of the value in terms of which they have identified themselves. They feel, possibly correctly, that there are many good things they would be giving up by acquiring the value in question and that, because the value is one with which they have (at most) only aspirational contact, they would be giving up these known goods for an unknown thing that may or may not be good.

It is, on the one hand, a testament to the flexibility of the human soul that even people who build a sense of identity around not valuing something (music, children, one's husband) can sometimes bring themselves to adopt the opposed value-perspective. People fiercely attached to independence can become wonderful parents. And couples who could not get their future into view without placing children at the center of it can struggle past the sadness of infertility to embrace the freedom of child-free life. It is, on the other hand, a testament to the difficulty of perspectival change that sometimes such people find themselves unable to see things differently, a fact that infuses the rest of their lives with a sense of loss. It seems possible to work to come to see a value one didn't see before, as well as to divest oneself of the value-perspective one currently has: this **(p.197)** is the process I have been calling aspiration. Aspirants have a more ambitious goal than self-maintenance. They work to build themselves up into something genuinely new.

On an aspirational account, self-creation is agent-driven learning in the domain of value. The aspirant brings herself to a different view as to what matters in life and comes to appreciate what she once did not. How does this contrast with self-cultivation? Consider an example with which Jean Hampton illustrates her "self-authorship" conception of aspiration:

[I]n the spring of 1991, American newspapers recounted the story of an investment banker who, as a teenager, wanted to be a clown. His parents strongly discouraged it, regarding it as inappropriate for someone of his background and abilities, so he went to MIT and got a job working in Silicon Valley in computers. Still he was dissatisfied and decided things might go better if he had an MBA. With this degree he got a job on Wall Street making a lot of money in a high-powered investment bank. But one day, he claimed, he woke up realizing that if he kept working on Wall Street, he would end up close to death never having gone to clown school. So he quit his job, and did exactly that. This is a nice story of someone who struggled to author himself, while under pressure to be what people in his social group expected of him . . . he faced pressure to submit to a social role, to take on preferences, interests, and projects that he did not really want. He experienced understandable relief when he reclaimed himself. (1993: 154–155)

The striking feature of Hampton's banker/clown, from an aspirational point of view, is how little learning he had to do. His challenge lay in resisting social pressures and exercising autonomy in the **(p.198)** service of doing what he really wanted all along. His self was already, as a teenager, developed in the direction of the core value of his life. He did not have to work to discover what he wanted; he had but to claim it, or later to "reclaim," the value that was present in him all along. Such a picture focuses our attention away from the developmental period during which one learns to see something in, e.g., clowning. I do not deny that one important way in which autonomy manifests itself is in the making of choices that align with who you really are, as opposed to who others want you to be. This form of autonomy is not available to everyone, however, since it presupposes the existence of a real or true self. On the aspirational model it can be true that you are at work on yourself, though there is, as yet, no real you; your wants are themselves a work in progress.

One way to see the difference between self-cultivation and aspiration is to look at the role allotted to secondary players. In Hampton's example, other people appear in the form of those who exert "social pressure" to abandon one's true desires. No doubt the man in question also had friends, family members, or associates who encouraged and perhaps facilitated his ultimate career change. One needn't, however, make mention of such people in order to tell this story of "self-authorship." He is in an important respect sufficient unto himself, equipped from the start of the story with the knowledge of what he wants. The aspirant, by contrast, reaches out to others for help in grasping what she wants. Tales of aspirational self-creation will, of course, feature the aspirant in a starring role, but they also lean heavily on a host of supporting players: teachers, mentors, (supportive!) parents, schools, advisers. Everyone relies on the care and love of the people around them, but aspirants rely on the people around them to care about and love the things they themselves are struggling to come to care about and love.

**(p.199)** Compare aspirational help with medical help. The patient doesn't know how to cure herself—on this point she lacks knowledge or ability or resources. But there is something she does know: she wants to be cured of the disease. In this kind of case, the agent's practical ignorance or inability is circumscribed by her practical knowledge or ability with respect to her desired end. She knows for certain that she wants a cure. It is not essential for her to acquire the knowledge or ability she lacks: she doesn't need to cure herself. In this sort of case, the person's ignorance/incompetence and her knowledge/competence can be cleanly separated from one another. The aspirant's ignorance, by contrast, runs "all the way down," in the sense that she cannot be sure, until she is no longer an aspirant, that she even wants to acquire the relevant value. And insofar as there is anything about the value she has not yet grasped, she sees it as her job to come to do so. She cannot farm this knowledge out to, e.g., a mentor in the way in which patients can farm medical knowledge out to their doctors. Aware that they have something to learn, aspirants lean on those more securely attached to the value in question, but this leaning is itself aspirational, displaying a pattern of ever-lessening pressure (see chapter 3, section II (c)). Like a child learning to ride a bike, the aspirant gradually replaces her helper's support with her own internal balance. This delicate interplay between one agent and others is one among the features of self-creation that is obscured if we assume the model of self-cultivation. And it is obscured, as I will argue, by the assumption that the created self (the endorsed or cultivated self) depends rationally on the creator self (the endorsing or cultivating self).

But how does the aspirational account of self-creation escape Strawson's dilemma?

**(p.200)** IV. Escaping Strawson's Dilemma

(a) *Normative Dependence*

Let me say that there is a *normative dependence relation* between two items when norms apply to the one item in virtue of the fact that, in the first instance, they apply to the other. When two things stand in this relation, the one inherits its normativity from the other. Let me illustrate with an example of a philosophical debate about the source of normativity.

We can assess intentions as good or bad, and we can assess actions as good or bad, and these forms of assessment do not seem to be fully independent of one another. Arguably, Aristotelians and Kantians agree that intentions and actions are subject to normative assessment, and they agree that these assessments stand in a relation of normative dependence, but they disagree as to the direction of dependence. A Kantian (or a Kantian of a certain stripe—it is not important for my point whether she is being true to Kant) says that intentions are the source of the applicability of norms to action. She contends that the fact that the intention has moral worth is what explains the normative status—the worth-having-ness—of the action. On her view, intentions contain a principle on which the agent conceives of himself as acting. The worth of the intention depends on the universalizability of this principle (or “maxim”); this worth then grounds the worth of the action the intention motivates.

Contrast this with a view on which it is actions that are, in the first instance, subject to norms of practical assessment. The Aristotelian opponent of my Kantian holds that actions are the proper objects of judgments of goodness and badness, i.e., praise and blame. The goodness or badness of actions grounds the goodness or badness of character, in that a good character is the character that disposes one **(p.201)** to perform good actions; character, in turn, grounds the goodness or badness of an intention, in that a good intention is whatever intention characteristically issues from a good character. Among other differences, these two theorists have a dispute about the location or source of moral worth: the one holds that actions have moral worth because (and insofar as) they are associated with certain kinds of intentions, and the other that intentions have moral worth because (and insofar as) they are associated with certain kinds of actions.

They agree that practical norms apply to both actions and intentions, but they have very different accounts of *how* they apply in each case. They have a dispute about the source of normativity. I am about to engage in just such a dispute with theorists of self-endorsement and self-cultivation.

(b) Priority of Created Self

Just as one can ask whether intention or action is the ultimate locus of moral worth, one can ask whether  $S_1$ , the creator self, or  $S_2$ , the created self, is the ultimate locus of value. Theorists of self-endorsement and theorists of self-cultivation agree in taking  $S_1$  as the normative ground of  $S_2$ . For the theorist of self-endorsement,  $S_1$  and  $S_2$  represent divisions within a person at a given time.  $S_1$  is one's evaluative system, and  $S_2$  is whatever feature (or prospective feature) of oneself one is using that system to evaluate. If  $S_2$  is endorsed, this is *because*  $S_1$  has done the endorsing. For the theorist of self-cultivation,  $S_1$  and  $S_2$  are different time slices of the person. If the person has cultivated a feature in herself, then  $S_2$  should have that feature *because*  $S_1$  did something—made a commitment or a resolution—picking that feature out as the one to be acquired.

The normative priority of  $S_1$  comes through particularly clearly in Joseph Raz's discussion of self-creation. Raz describes  $S_1$  as **(p.202)** having done something akin to making a promise, and  $S_2$  as therefore having a new reason for action that is akin to one's reason to keep a promise. Raz says:



Our life comprises the pursuit of various goals, and that means that it is sensitive to our past. Having embraced certain goals and commitments we create new ways of succeeding and new ways of failing. In embracing goals and commitments, in coming to care about one thing or another, one progressively gives shape to one's life, determines what would count as a successful life and what would be a failure. One creates values, generates, through one's developing commitments and pursuits, reasons which transcend the reasons one had for undertaking one's commitments and pursuits. In that way a person's life is (in part) of his own making. It is a normative creation, a creation of new values and reasons. It is the way our past forms the reasons which apply to us at present. But it is not like the change of reasons which is occasioned by loss of strength through age, or the absence of money due to past extravagances. Rather it is like the change occasioned by promising: a creation, in that case, of a duty one did not have before. For, whatever reasons one had to make the promise, its making transforms one's reasons, creating a new reason not previously there. Similarly, the fact that one embraced goals and pursuits and has come to care about certain relationships and projects is a change not in the physical or mental circumstance in which one finds oneself, but in one's normative situation. It is the creation of one's life through the creation of reasons. (1988: 387)

In Razian self-creation,  $S_1$  holds the reins, creating values and reasons for  $S_2$ . Raz responds to the objection that such creation is **(p.203)** arbitrary by leaning on the analogy with promising: “It may have been wrong to promise to give my son fireworks, for they are too dangerous. But having made the promise it may now be my duty to give him the fireworks” (388). This analogy plays right into Strawson’s hands: *Why* did the father make that promise (i.e., create those values)? Had a yet earlier self promised to make these promises? Or did he promise for no reason, arbitrarily? Raz’s vulnerability to Strawson’s dilemma is, I propose, a product of his understanding of the normative grounding relation between  $S_1$  and  $S_2$ . Promising exemplifies a form of agency in which the earlier self is normatively prior to the later one, and Raz takes this structure to be present in self-development: one’s earlier self makes it the case that certain norms govern one’s later self.

Let us review the dialectic thus far. We began by considering whether someone could create herself by endorsing or affirming some (prospective) value. We found that if a person stepped back far enough to avoid simply reaffirming what she already values, her endorsements no longer constituted value-generation. The self that is in a position to fairly, impartially evaluate its own values issues not values but, at best, judgments as to rational entailments among values. But could, perhaps, a judgment as to what one’s current set of values entails form an initial stage in the process of self-creation? From this question emerges the self-cultivation model, which describes the process of following through on such a judgment by making oneself into the person one judged one ought to be. But self-cultivation is vulnerable to Strawson’s dilemma, in response to which we found ourselves in a regress, stepping backwards in an attempt to identify ever-earlier selves from which an act of self-creation might (non-arbitrarily) spring.

The problems seem to emerge from the fact that we press each model backwards, asking, “Who endorses the endorser?” and “Who **(p.204)** cultivates the cultivator?” This backward pressure tracks the direction in which normativity is, as it were, flowing. Both models make  $S_1$  the source of  $S_2$ ’s normativity. The regressive line of questioning seeks  $S_1$ ’s normative advantage over  $S_2$ , which is to say, the source of  $S_1$ ’s authority to endorse, or make commitments on behalf of,  $S_2$ .

It is important to note that Strawson's dilemma does not indicate a problem with the very idea of self-endorsement or self-cultivation. The theorist of self-endorsement can explain why the evaluating self typically has normative authority over the self it evaluates: the former simply *is* the locus of value. It is in the nature of one's values that they are in a position to evaluatively endorse or reject our desires, motives, habits, etc.—but not our values. For those items are *members* of  $S_1$ . Only in the case of value-endorsement does the problematic demand to “step back” from our values find application.

Likewise, self-cultivation is an intelligible way to make sense of efforts at self-change. Suppose  $S_1$  concludes, after careful reflection, that he ought to quit smoking. Holton can explain why this resolution not to smoke binds  $S_2$  when the temptation to smoke strikes.  $S_2$  has reason to enact—without reconsidering the issue—the judgment that he made under rationally favorable conditions. We form resolutions (in part) because we anticipate the phenomenon Holton calls “judgment-shift,” in which occurrent desire sways our judgment as to what we should do.  $S_1$  was in a better position to adjudicate  $S_2$ 's decision than  $S_2$  is, so it makes sense that norms should apply to  $S_2$  in virtue of  $S_1$ 's agency. The problem is that  $S_1$  has no such advantage over  $S_2$  in the case of value-creation.

In the case of value-creation, there is no special feature of  $S_1$  that explains *why* she should be authoritative over  $S_2$ . For this reason, the theorist of self-cultivation is left grounding  $S_1$ 's authority in the act of cultivation itself:  $S_1$ 's authority lies simply in the fact that she came first. But a vicious regress results from explaining the **(p.205)** cultivating self's rule-making authority by appeal to her status as cultivator, for there are other, even earlier selves. Thus  $S_1$  will have to derive her normative license from a yet earlier cultivator, and we will be hurled backwards toward an ever-receding source of ultimate authority. I propose that this regress reflects the fact that, in the case of self-creation,  $S_1$  cannot do the work that self-endorsement or self-cultivation would require of her. She does not have the normativity she is supposed to bestow.

I submit that the theorist of self-creation needs to get the creator self looking forward rather than backward: instead of imagining my future self as beholden to my past self, I suggest we imagine my past self as looking forward, trying to live up to the person she hopes to become. The creator self doesn't *make* a promise; she sees (to take up another facet of the concept of a promise) a *promise* of a better self. When we speak of some prospect as a promising one, we do not use the word "promise" literally since, among other reasons, what doesn't exist yet cannot make promises. But this suggestive locution captures the reversal I propose. Promising presupposes a certain stability and predictability in one's self and one's circumstances. In a case where the values in question are in need of being created, it is only  $S_2$  and not  $S_1$  who would be in a position to take on such commitments. It is not  $S_1$ 's place to embrace goals on behalf of, form reasons for, or create duties that will bind the created self. By her own reckoning  $S_1$  is not as she evaluatively should be.  $S_1$  is not the lender but the borrower of normative authority.

### (c) Strawson's Two Requirements

We can restate Strawson's dilemma in terms of the idea of normative dependence. The one horn—no random changes—calls for a normative dependence relation between the two selves, while the **(p.206)** other—no derivable values—demands novelty in the created self. We can represent these demands thus:

*The Continuity Requirement* says that  $S_1$  and  $S_2$  stand in some normative dependence relation.

*The Novelty Requirement* says that  $S_2$  must contain a value or values not dependent on the values of  $S_1$ .

The continuity requirement and the novelty requirement contradict one another only if we fill out the former by assuming that  $S_1$ 's values are the source of  $S_2$ 's. In the next section, I want to flesh out what abandoning this assumption looks like. We will find that when we allow for the possibility that a process of value-acquisition progresses toward its own source of normativity, Strawson's dichotomy gives way to a third option: that the creator self relate to the self she creates aspirationally.

### (d) Self-Creation by Aspiration

Consider how people come to appreciate the value of classical music or religious observance or fashionable dress or fine cuisine or political debate. The process might begin because one is suddenly, experientially confronted with a value one's prior valuations did not lead one to anticipate. One might experience such a transformative moment at a performance or in a church or among new friends. Such experiences are "transformative" only in the inchoative sense. They will not bear the weight we will have to put on them, if we try to use them to extrapolate a commitment to the endpoint. The transformed agent will look back, years later, and say of that initial event, "Little did I know, back then, what was really valuable about classical music" (or religious observance or fashion, etc.). She can, at that later point, say, "*This* was what I was after all along," but only **(p.207)** because she encountered "this" before knowing exactly what "this" was. The value she comes to endorse is one she knows as a result of working toward a target she could not, at the time, exactly envision.

It is undeniable that many of our values trace their roots to that early-childhood period in which our ethical development was managed by others. Perhaps the Strawsonian will want to say, with respect to these values, that we simply find ourselves with interests and mastery we played no role in generating. And then he will tell us that we can work up additional values by rationally developing the commitments implicit in them. Alternatively, we can once again be passively subject to environmental influence and acquire values that bear no rational relation to those already inculcated in us. But are these the only choices? For all of us have developed passions—for fine food, politics, music, or philosophy—long after leaving our parents' homes. And if we consider the course of such development, we find, I submit, that the Strawsonian picture simply doesn't ring true.

Given the expertise and work involved, it is implausible that anything but the earliest stages of such a transformation can be explained through fully external factors. For instance, the fact that someone found himself, for incidental reasons, in the exceptional gastronomic environment of Osaka, Japan, might be the beginning of the story. Those experiences could ignite a spark of interest, but then something more would be needed to drive someone's systematic development of that initial spark into a full-fledged passion. The "something more" in question is unlikely to be a value to which he was antecedently committed, from which a passionate interest in culinary excellence could be derived.

I don't want to deny that there might be cases in which an accidental and transformative initial experience or an anterior value-commitment suffices to explain someone's value-transition. But in **(p.208)** many cases, we embark on these sorts of adventures without thinking that we know, in advance, exactly what we will get at the end. And that is not to say that we take them up on a whim, or for no reason, or by accident. There is an intermediate possibility that Strawson's dilemma has directed us away from recognizing, namely the possibility that someone has an inkling of a value he does not fully grasp. He doesn't have a fully worked out sense of how this value fits into the rest of his values, because he doesn't have a fully worked out sense of what this value is. How could he, if the value corresponds to the intrinsic pleasures of the fine discriminations he is not yet capable of making? Nor is his pursuit conditioned on the coherence between the new value and the rest of what he currently values. Indeed, his ardent pursuit of it may take him away from much of what he currently cares about.

Most of the profoundly important activities, relationships, and forms of knowledge that human beings pursue are ones a person can fully appreciate, and integrate into her value-system, only once she is well acquainted with them. And our question concerns the process of becoming acquainted. If you had to acquire values either by accident or by working out the entailments of your prior commitments, there would be no such process. You would either already be, in effect, at your value-destination or have no way to get started. But those don't seem to be our only options.

The way in which people stand toward many of the values that they do not fully appreciate is that they partly appreciate them. And with respect to some of these partly appreciated values, they also have the inclination to appreciate them more. They have a sense that their inchoate appreciation is *incomplete*, and act in order to attain a better valuation-condition. The actions they perform are versions of the actions they will be in a position to perform once they have fully acquired the value: i.e., the one who wants to value music acts **(p.209)** like the person who already loves music. But she also acts *unlike* the person who already loves music, since that person is not herself acting like anyone else. Such people are, in effect, imitating or trying to live up to someone. They don't pre-approve of the person that they're trying to be; rather, they hope that the person they aspire to be would (and will!) approve of them. They see themselves as the imperfect version of that person who, in turn, serves as the standard by which they are to be assessed.

If you are trying to get better acquainted with some value, then you take your antecedent conception of that value to be inadequate. You act in order to grasp the value better, but your reason for wanting to grasp that value must be the very value you don't yet fully grasp. Life is full of moments in which one contemplates some obscure value from a great distance. We can't comprehend the value of child raising for us, let alone the value of the life of the child we will raise, before starting a family. We go to college for the education college will itself teach us to appreciate; we leave our hometown with the aim of making some foreign place home; we date in order to love, and get married in order to love in a new way; we choose a career because of the as yet unfamiliar joys of expertly doing the work in question. In pursuing these values, our attitude is not merely a hope or wish that we will one day come to appreciate them. We *work* to appreciate them, and this work is rationalized and guided by the values we are coming to know. In these cases, the full justification of what we are doing can come only at the end of the story. It is the end that provides the normative standards for assessing what comes before it.

(e) An Objection: The Value Gambit

Why think of the aspirant as being under the guidance of her future self, as opposed to being under the guidance of the very value to **(p.210)** which her future self is (more) responsive?<sup>7</sup> The attraction of this alternative proposal lies in the prospect of avoiding an invocation of teleology. For if the value of, e.g., classical music guides one's aspiration to appreciate it, and if we can suppose that the value pre-exists the agent's arrival at her aspirational goal, then the source of normativity needn't *follow* that of which it is the source. I want to explain why this gambit does not succeed.

The proponent of the value gambit must acknowledge that the aspirant is guided by the intrinsic value of, e.g., classical music in a different way than the person she aspires to become. The latter's response to the value is indicated by, e.g., the joy she takes in the music she listens to, her articulate sensitivity to different renditions of a piece, her regular subscription for season tickets, her disappointment when she must miss a performance, her delight in sharing her love of music with her children. The aspirant, by contrast, is moved by the value of classical music to seek the approval of her music teacher, to commit to meeting a friend at the symphony so that she will not back out and see a movie instead, to pinch herself to stay awake throughout the piece, etc. If we want to characterize the aspirant as guided by the value of classical music, we will have to allow that she is guided in a non-paradigmatic way.



This non-paradigmatic form of value-guidance must be understood with reference to the full-fledged value-guidance into which it aims to develop. Pinching yourself to stay awake isn't a mark of appreciating classical music; it's a mark of *not* appreciating classical music. It is a form of response that comes under the aegis of the value only to the extent that we situate it in a developmental process: pinching in order to eventually attend **(p.211)** without pinching. It is only insofar as we bring the aspirant's response into relation with the response of the person she aspires to be that we see the former response as guided by the relevant value. The sense in which it is the value of music that guides her to care about her teacher's approval or to pinch herself is that it is her (ultimate) appreciation of the intrinsic value of music that makes intelligible what she is now doing—including the way in which she is now being guided. The value of classical music makes it right for her to care about her music teacher's approval or to pinch herself to stay awake, but this is because—and only because—she does those things so as to (really, fully) be guided by the value of music.

Aspirationally responding to a value isn't on all fours with non-aspirational response to it, in the way that, e.g., one can respond to musical value both by attending and by discussing. Rather, aspirational response to the value becomes legible as a response to the value only in the light of the proper response into which it develops. And this proper, intelligibility-conferring response is not one the aspirant herself grasps—except, of course, aspirationally. If this is correct, the shift from self to value fails to avoid reference to teleology. If we say that the aspirant is guided by value in a way that is derivative of her later mode of guidance, we adopt a form of explanation no less teleological than the one needed to claim that *S2* is the source of normativity. In either case, a later instance of normativity renders intelligible the normativity present earlier.

It might, nonetheless, seem to make a big difference whether what a person is guided by is her self or value. For the first characterization might strike a person as objectionably egoistic. After all, what if someone aspires to be humble or kind—can these aspirations really be understood as forms of concern with *oneself*? I think, first, that we must recall that we are analyzing the self in terms of **(p.212)** value.<sup>8</sup> When someone aspires to be the person she will become, she is aspiring to a form of value-appreciation; she is looking to fully inhabit a value. Given the theory of the self we are taking for granted, it doesn't matter much whether we adopt the language of self or value. For even if we choose the latter, we will have to allow that the defective way in which the aspirant responds to the value places a distorting emphasis on her own acquisition and ownership of the value. Whether we speak of her as guided by her future self or (aspirationally) by value, we will have to acknowledge that we have depicted the aspirant as more self-oriented than she should be. But this is a feature of the theory, not a bug: aspirants *are* more self-oriented than they should be!

Because an aspirant always has an eye on her own progress—she is engaged with trying to become someone—she cannot be fully, properly responsive to, e.g., music or the feelings of others. The would-be music-lover's attention must be divided: some of it can be directed at the music, but some of it must be set aside for considering the question of how much attention she is paying to the music (“Am I letting myself get distracted?”). When the kindness-aspirant acts unkindly, her ability to appreciate the suffering she has caused is impeded by worries over her own kindness-trajectory (“I've been working so hard on empathy!”). It is no paradox that a kindness-aspirant can't really, fully be kind. She is, after all, *not* fully kind, which is why she is aspiring to be.

### V. Teleology and Agency

#### (a) Backward Causation?

The aspirational account of self-creation offers us a recognizable story of what it is like to actively acquire values and avoids the **(p.213)** paradoxes and problems associated with trying to force our understanding of that transition into the mold of self-endorsement or self-cultivation. It may nonetheless arouse suspicions due to its teleological character. We are not, generally, inclined to accept forms of explanation in which that which is later grounds what is earlier.

The opponent of teleology may suspect that I am invoking something like backward causation or time travel: the future cannot cause the past, so how can something that is *in* the future have an effect on what I am doing *now*? Unless the ground of the agent's self-transformation is there at the outset, how is she to move herself to the desired result? In response to this worry, I want, first, to distinguish the claim that  $S_2$  is a normative source of  $S_1$  from the absurd claim that  $S_2$  causes  $S_1$  to come into existence.

When an agent acts on a proleptic reason, she takes steps toward bringing about the normative source of her present condition. She works to make herself grasp the reason ( $R_2$ ) that is, normatively speaking, the full or complete or nondefective variant of the proleptic reason causally driving that work ( $R_2$ ). Her grasp of  $R_1$  is causally responsible for her grasp of  $R_2$ , though  $R_2$  is the source of normativity for  $R_1$ — $R_2$  is what makes  $R_1$  a *good reason*. My claim, then, is that in the case of aspiration, causal or temporal priority fails to track normative priority. Is this possible? I am inclined to think it must be, since it is actual. The theorist of aspiration gives us what is the only way to account for self-creation in the face of Strawson's dilemma; and self-creation is a fact of life. We ought, therefore, to accept that such a reversal is possible, unless someone can show otherwise. One person who might take herself to be in a position to offer such a demonstration is the theorist of action.

Christine Korsgaard has articulated what we might call a "guidance condition" on rational agency: "A person acts rationally . . . only when her action is the expression of her own mental (**p.214**) activity" (1996: 33); "A rational agent would be guided by reason in the choice of her actions." The action theorist may argue that in order for some behavior to be assessable as succeeding in a distinctively practically rational way, the norm with reference to which she is supposed to have succeeded must be mentally present to her at the time of action. For only in that way could the norm have guided her action. In order to understand aspiration as the aspirant's own work, we need to be able to understand it as coming *from her*.

I propose to set aside more general worries about teleology and focus attention on the worry specific to the action theorist: Can we represent the aspirant's progress as something she does, given that she lacks a clear or determinate conception of her target until she arrived at it? Does the aspirant satisfy Korsgaard's guidance condition?

### (b) Two Conceptions of Agency

If Davidson's causal theory of action, read a certain way, is correct, she will not. Suppose that some behavior counts as an example of agency in virtue of the fact that it was caused by a belief and desire that rationalize it. On this picture, rational guidance is a causal matter. For instance, my trip to the store is caused by the desire to get food and the belief that I can get food at the store. My trip succeeds if I do indeed end up getting food at the store, and it can be assessed in the light of this norm precisely because this norm played a causal role in the genesis of the trip. On at least one way of understanding the desire that functions in such a causal explanation, it must constitute a grasp of what outcome is sought. Some event counts as agency to the extent that its success conditions are inscribed in the attitudes causing it. If practically rational guidance required an agent to know exactly what she wanted out of the outcome, aspiration **(p.215)** could not qualify as rationally guided. The aspirant fails to grasp the full normative grounding of her project until it is completed. And if the reason is fully available to the agent only at the conclusion of the aspiration, it cannot (efficiently) cause the behavior that constitutes her aspiration.

A number of philosophers have recently developed a line of thought from Anscombe into a rejection of the causal theory of action. The alternative can be expressed as a response to the well-known problem of "deviant causal chains." The worry is that the Davidsonian has no way of ensuring that it is really the reason, as opposed to something correlated with the presence of the reason, that is doing the relevant causal work.<sup>9</sup> For instance:

A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally. (Davidson [1973] 1980: 79)

Davidson says that he himself “despair[s] of spelling out . . . the way in which attitudes must cause actions if they are to rationalize the action” (79). Some, perhaps ultimately including Davidson himself,<sup>10</sup> have taken the problem as fatal to the causal theory. If we **(p.216)** cannot eliminate wrong causal chains, then we may be inclined to give up on the project of basing our answer to the question of whether some event is an action on the way in which that event was caused.

Some of Davidson’s critics<sup>11</sup> have sought refuge from this problem in Elizabeth Anscombe’s *Intention*. According to the account they find in Anscombe, some behavior counts as an action—and therefore as assessable in the light of some end—not in virtue of the beliefs and desires that constitute its cause, but rather in virtue of the fact that, throughout the course of the action, the agent has a distinctive practical knowledge of what she is doing. So some behavior counts as, e.g., a hand-raising done for the sake of voting, because the agent does it in the practical knowledge that that is precisely what she is doing. This kind of knowledge is distinctive in being non-observational and non-inferential. When *I* try to answer the question as to what she is doing, I must observe her behavior and perceive that her hand is going up. In order to know that she is raising her hand in order to vote on the proposal, I must draw an inference from that (and other) observations. When *she* tries to answer the question as to what she is doing, she needn’t observe herself or draw any inference. Thus observational or inferential knowledge of what someone is doing contrasts with the immediate, first-personal knowledge characteristic of the agent acting. Does the Anscombean account of agency offer better resources than the Davidsonian theory to the theorist who is accounting for aspirational agency?

**(p.217)** It does not. If the agent's reason is to be the object of her practical knowledge (or even her practical belief), she must fully grasp it in advance of the action's coming to an end. And this is just what the aspirant cannot do. She does not know, or even take herself to know, what she is doing. She is dissatisfied by her own answer to the "why?" question. Whether we make the agent's reason for action the efficient cause of the action or the object of the agent's practical knowledge, we presuppose access to the normative standards governing one's own action. But the aspirant lacks this access, being unable to fully articulate, either before or during her action, exactly what she is doing and why. This is not to say that she has *no* idea what she is doing, but rather that her conception of what she is doing is derivative of a superior conception she will have after she has arrived at her destination.

It will be helpful to contrast aspirational ignorance of the end with the merely general or schematic grasp of an end to which specificationists such as David Wiggins (1975) and Henry Richardson (1994) have called attention.<sup>12</sup> Sometimes an agent's grasp of an end is too abstract to allow her to reason instrumentally in its service; she must, first, figure out in more detail what it is she seeks to do. For instance, if someone would like to have an entertaining evening,<sup>13</sup> he needs to figure out what kind of entertainment (i.e., exciting vs. relaxing, social vs. private) before he can begin to reason about how to arrange it. If a politician would like to create a point-to-point, no-transfer public transportation system for Paris, she needs to make the idea more concrete before she can turn it over to engineers. If you want to do something big for your wife's **(p.218)** birthday, you must settle whether what form this will take (romantic getaway for the two of you, or party for friends and family, or extravagant gift, or day to herself) before you can begin making the arrangements.

A person who has not yet done the relevant kind of specifying could be said not to know (exactly) what it is for the sake of which she will act and to reason in the service of filling out the end rather than finding a means to it.<sup>14</sup> How does her ignorance differ from the aspirant's ignorance? The distinction between general and particular, on which the specificationist relies, is different from (and cuts across) the distinction between the aspirant's partial or inchoate grasp of value and the full grasp to which she aspires. The former distinction allows one to factor out a known (general) element and an unknown (specific) element, facilitating action that fulfills the end to the extent that it is known. For instance, if I want to have an entertaining evening, I may know that I need to finish my work this afternoon. Without knowing whether my evening will be relaxing or exciting, I know that I won't be able to enjoy it with work hanging over my head. My schematic grasp of my goal may be perfectly adequate to dictate certain means taken in its service. Both my knowledge (that I want to have an entertaining evening) and the action (finishing up my work) that it prompts are shielded from my ignorance (as to what form that entertainment will take). I may not know what I will do, but I know why I am doing what I am doing.<sup>15</sup>

**(p.219)** In these cases, the ignorance of the end doesn't touch action-explanation. Indeed, there is no reason why those who engage in specificationist deliberation need exhibit any ignorance of their end at the time of action. For in many cases nothing prevents them from holding off on acting in the service of their end until they have completed the relevant course of deliberative specification; and in some cases, they will not be able to act until they have done so. (Of course, there may, at that point, still be uncertainties as to the means by which the end will be realized.) It is, by contrast, characteristic of the aspirant that she *must* act in ignorance of what she is doing, since it is by such action that she comes to learn both the value and the nature of her activity. When the specificationist's agent acts, her action is directed at satisfying or realizing her end, as far as she already grasps it, not at learning what that end is. One place to see the difference from aspiration is in each agent's self-understanding as regards the assessment of her action. Aspirants can't confidently and authoritatively assess their own actions. They cannot tell whether they are doing what they are doing well; submit themselves to the assessment of others. Agents with a schematic grasp of their end do not betray a similar reliance on the assessments of others.<sup>16</sup>



The agent who has a schematic grasp of her end acts in a way that can be fully explained by the grasp she already has, however schematic it is. The aspirant acts in a way that can be explained only by the grasp she will have when her agency is complete. The aspirational target cannot be read off the agent's antecedent beliefs and desires unless those beliefs and desires are, in turn, interpreted proleptically, i.e., in relation to what comes after them. The end for the sake of which the agent acts when she aspires is not itself to be assessed in the light of what she currently takes herself to be doing or the desires with which **(p.220)** she entered into the pursuit. Rather, the order is reversed: both the intentions with which she acts and the motivations from which those intentions spring are to be assessed in terms of their conduciveness to her aspirational end. Thus we cannot say, of aspirational agency, that the normative assessability of its results derives from the fact that some conception of those results was present to the agent during the course of the agency. In the case of aspiration, the representation in question will have to be assessed in the light of another representation, namely the resultant one. If we want to ground the normativity of aspiration in antecedent representation, we will run in a circle.

We could accommodate aspiration within Davidsonian or Anscombean action theory by breaking it up into its many component actions. We might present each of those actions as caused by, or done with the knowledge of, some reason that rationalized it independently of the value that stands at the end of the process as a whole. On such a piecemeal picture, the many actions add up to a change of value at the endpoint, but that change of value is not a target of her agency. Indeed, the proximate face of someone's proleptic reason will provide us materials for this sort of reduction: we can say that I listened to the opera because I promised myself a chocolate bar when I got to the end. But we have seen, in chapter 2, that such a picture of aspiration will not do: the theorist of aspiration must supply a rationalization of the process of aspiration as a whole. Treating each action in isolation from the ultimate aspirational target produces a distortion of what the agent is doing; on the right way of looking at these moments of agency, they are not self-standing.

(c) Aspirational Agency versus Self-Standing Action

It is, perhaps, no flaw in the Anscombean or Davidsonian account that they do not accommodate aspirational agency. They were, after **(p.221)** all, engaged in action theory as opposed to the more general discipline one might call agency theory. And though some actions (e.g., building a cathedral) may take a long time, it does not seem natural to characterize the whole process of becoming a music-lover or a parent as a single action. It is easier to apply that label to smaller projects—such as taking a music class or outfitting the nursery. But if I am right, those actions will, insofar as they form part of aspiration, be of a distinctive kind: they will be actions that cannot be understood except with reference to a larger stretch of agency of which they are a part. Anscombe and Davidson were offering accounts of what we can now characterize as “self-standing actions.” The aspirant’s agency does not have this quality: her individual actions are not self-standing, and the aspirational whole of which they are a part is not an action. Anscombe and Davidson described a particular kind of agency, namely the self-standing agency of the clear-eyed agent. The theorist of aspiration corrects them only by adding that there is another form of agency, one appropriate to the practical learner.

Let me return to Korsgaard’s guidance condition: “A person acts rationally . . . only when her action is the expression of her own mental activity”; “A rational agent would be guided by reason in the choice of her actions.” We can now see that statements such as these contain a crucial ambiguity. Suppose we paraphrase the condition at which they gesture as follows. In order for some stretch of agency done at time  $t$  to be done for the sake of norm  $N$ , it must be true that the agent grasps  $N$  at  $t$ . What I hope to have illustrated is that there are two ways of grasping a norm (or being guided by reason or having one’s action be the expression of one’s mental activity). One way is in the manner of the agent of a self-standing action. Such agents’ activity is grounded in the bedrock of their current grasp of norms: in Anscombean terms, they act from practical knowledge. **(p.222)** But it is also possible to grasp a norm proleptically. In such a case, the agent’s grasp of the norm guides aspirational activity without serving as its normative bedrock, for it is but an attenuated version of the grasp of value she will have once she reaches her aspirational endpoint. *That* grasp constitutes the normative bedrock, the ground on which all of her agency up to that point rests.

The practical reasoners I am describing seek to acquire what Anscombe's agents are fortunate enough to be in a position to exercise. Instead of acting, at each moment in time, *from* practical knowledge, the thread that binds together the various things they do (over the course of months, or even years) is the knowledge *toward* which they act. They are not knowers, but learners. Proleptic rationality illustrates the possibility that we might engage in practical reasoning in order to improve the very conception of the good that drives that reasoning.

Most instances of agency are teleological, in that they are done for the sake of some subsequent end. The agent's behavior is *going somewhere*, and that is because it is guided, from within, by the agent's sense of where she is going. What makes instances of aspirational agency special—puzzling—is that they are cases in which the agent's sense of where she is going is also, in the relevant sense, *going somewhere*.

Notes:

(1.) *Beyond Good and Evil*, sec. 21. Are these skeptics right to see Nietzsche as an ally? He is, after all, apt to wax eloquent on the power of the strong to create themselves. See the passages cited in Owen and Ridley (2003) in response to the more straightforwardly fatalist reading offered in Leiter (2001).

(2.) Laura Ekstrom (1993) modifies Frankfurt's account by adding such considerations of coherence to it. She aims thereby to defend Frankfurt against Watson's charge that any higher-order desire must itself be endorsed by a yet higher-order desire, generating a regress. On Ekstrom's picture, my desire is "mine" because I "authorize" it in virtue of its occupying a position of coherence among my desire set: "The coherent elements 'fit with' the other items one accepts and prefers, so that, in acting upon them, one is not conflicted" (609). Ekstrom's defense suffers from a systematic ambiguity as to whether what makes a desire mine is the fact that it coheres with the rest of my desires or the fact that I take it to cohere with them and therefore authorize it: endorsement and coherence compete for centrality in her account. Even if it worked, however, her defense could not go so far as to offer a theory of self-creation. For Ekstrom identifies a person's values with those desires that "fit with" one another ("I propose to define coherence with one's character system . . . as determined by what it is valuable to pursue within that system"; 610), and this means that one cannot so much as articulate the question as to which values, i.e. which (among many) coherent system of desires, one should "authorize."

(3.) Supposing, for the sake of argument, that it really is rational to become more coherent. For an argument against the existence of demands of this kind, see Kolodny (2005, 2008).

(4.) One may worry that this argument generalizes in a problematic way: Isn't there such a thing as *cognitive* coherentist self-policing, in which one assesses one's beliefs for coherence? If we assume it is possible to generate and destroy beliefs by examining oneself for coherence, that is because we take the cognitive case to have the following peculiarity: judging that one ought to believe that *p* is or entails believing that *p*. This rule may or may not hold for believing, but it certainly does not hold for valuing. For valuing, as argued earlier, has not only a cognitive but also a motivational and affective element. The last feature is worth singling out as particularly recalcitrant: judging that one ought to feel something does not even tend to give rise to the relevant feeling.

(5.) Strawson's ultimate target is responsibility for action: he wants to show that because we cannot create our values, we cannot be responsible for anything we do from those values. I discuss the major premise—that action-responsibility derives from character-responsibility— in chapter 6 part V. Since he does not always clearly separate the arguments for these two claims and since I am not discussing moral responsibility, what I offer here is not a direct paraphrase of any passage of his paper. Nonetheless, both the style and the substance of my presentation of the dilemma are heavily indebted to Strawson's way of framing (what he takes to be) an objection to both self-creation and moral responsibility.

(6.) See Noggle (2005: 108 n. 26); McDowell (1998: 36–37).

(7.) Thanks to Sam Scheffler and Matt Boyle for pressing this objection.

(8.) See Introduction, part IV.

(9.) This well-known worry can be articulated negatively, as to how one rules out deviant causal chains (Frankfurt 1978: 157–158; Davidson [1973] 1980: 78) or, positively, as to whether the semantic element is doing the causal work (Dretske 1989). For a recent defense of Davidson on this point against a rival neo-Anscombean picture, see Paul (2011).

(10.) “Several clever philosophers have tried to show how to eliminate the deviant causal chains, but I remain convinced that the concepts of event, cause and intention are inadequate to account for intentional action” (Davidson 2004: 106).

(11.) Here I rely on Paul (2011), which draws together work by Michael Thompson, Sebastian Rödl, Candace Vogler, Richard Moran, Martin Stone, Matthew Boyle, Douglas Lavin, Kieran Setiya, and J. David Velleman. She helpfully abstracts the differences between their views to isolate an Anscombean emphasis on the formal, as opposed to efficient, causal structure of action (cf. Paul 2011: 5 n. 12).

(12.) I thank Martha Nussbaum and Elijah Millgram for suggesting that my position be clarified by the contrast with specificationism.

(13.) This example is from Williams ([1980] 1981: 04); the next one is from Millgram (2008), whose discussion of specificationism informs the presentation of it offered here.

(14.) Though Candace Vogler (2002: 159–169) has argued that specificationist reasoning is a species of instrumental reasoning.

(15.) When the aspirant buys a bus pass to get to her music class, the purchase inherits her uncertainty about the end: if you press her, you will be able to get her to admit that she doesn't really know why she's buying the bus pass, since she doesn't really know why she's taking the class. What she doesn't know infects every element of the aspirant's action. I thank Anton Ford for helpful discussion of this material.

(16.) I thank Gabriel Lear for this point.



Access brought to you by: