

## CHAPTER FIVE

### A PROPER ACCOUNT OF THE RELATIONSHIP BETWEEN SELF- CONCEPTIONS AND PRACTICAL REASONS

#### 5.1 Introduction

The goal of the preceding chapters has been to survey views that bear on the relationship between self-conceptions and normative reasons. The goal of this chapter will be to parlay this information into a positive account of this relationship. To this end, I have three specific aims in the pages that follow. First, I will review Korsgaard's and Frankfurt's accounts of this relationship to determine what they got right and what they got wrong. Second, I will construct my own account of the self-conceptions/normative reasons relationship. Third, I will enumerate some practical principles implied by this account that provide formal guidance on how to approach self-conception related questions in one's deliberations. By this chapter's end, I will have completed the first of my project's three central goals, which is to give a proper account of this relationship. In so doing, I will have also set the stage for my project's next central goal—i.e. accounting for why there appears to be a relationship here in the first place—which will serve as the focus of the next chapter.

The argument for the version of this relationship that I will be defending comes in two distinct stages. The first stage, which takes as its starting point Frankfurt's Co-Determination Thesis (CDT), corrects for deficiencies that are “internal” to Korsgaard's and Frankfurt's accounts (i.e. deficiencies that are unrelated to Chapter 4-based objections to their accounts). The output of this first stage will be a “subjective” first approximation of the view that I will be defending, the Modified Co-Determination Thesis (MCDT). The second stage applies the views defended in Chapter 4 (concerning the relationship between normative reasons, values, and facts) to this subjective version of MCDT. The reason for this two-stage approach is ease of

exposition; given the number of considerations that figure into the CDT-to-MCDT transition, our task of understanding how we get from the first view to the second will be made easier if we treat these two categories of considerations separately.

In terms of how this discussion will proceed, in Section 5.2 and 5.3 I evaluate Korsgaard's and Frankfurt's accounts of this relationship, ignoring for the moment possible Chapter 4 objections to them. Section 5.4 presents a first approximation of MCDT that corrects for deficiencies in their accounts. In Section 5.5, I consider how the two central theses of Chapter 4—the Value-Based Reasons Thesis (VBR) and the Fact-Based Reasons Thesis (FBR)—count against Korsgaard's and Frankfurt's views, and then in Section 5.6 I apply VBR and FBR to my earlier, first approximation of MCDT. I close in Section 5.7 by enumerating some practical principles that are implied by this discussion.

## **5.2 Korsgaard on the Relationship Between Self-Conceptions and Practical Reasons (Revisited)**

Let us begin by looking back at Korsgaard's account of this relationship. It might be useful to recall the first six steps of her argument, which run as follows:

1. Reflective distance from my desires gives me the inescapable need to act on the basis of reasons (if I am to act at all).
2. Reflective distance from my desires also forces me to have some self-conception as (at least) the thing that is distinct from my desires.
- \* [Korsgaard's Expression Claim: My normative reasons are expressive of my self-conceptions.]
3. My normative reasons are grounded in my self-conceptions.
4. My normative reasons are grounded in objects of my concern (things that I desire or care about).

5. A practical identity is a self-conception that the agent cares about.
6. My normative reasons for action are grounded in my practical identities.

The upshot of Korsgaard's discussion is that all of my normative reasons for action are determined by my practical identities. Every time I have a normative reason to act, this reason exists because I hold an associated, endorsed self-conception that gives my life meaning. I have normative reason to care for my wife, for example, because I hold the practical identity "Nicole's Husband," which gives my life meaning. As we saw in Chapter 2, Korsgaard's account of this relationship is unidirectional; endorsed self-conceptions ground normative practical reasons, but not vice versa.

The first question to ask is what we can learn from Korsgaard's account. What did she get right? While I will be quite critical of the particulars of Korsgaard's account, at the very least she deserves credit for being one of the first philosophers to discuss in some detail the relationship between the content of self-conceptions and normative reasons. Not only is Korsgaard sensitive to some rather interesting intuitions concerning this relationship, she was arguably the first to attempt to build a systematic account of it. In these respects, she has managed to bring attention to an under-discussed topic of some importance.

This brings us to the question of what Korsgaard got wrong.<sup>1</sup> One rather obvious concern involves her commitment to DBR, but since this issue will serve as the focus of Section 5.5, I will refrain from addressing it here. Focusing for the moment on worries unrelated to this commitment, the most pressing objection

---

<sup>1</sup> In discussing only the value of Korsgaard's appreciation of this issue, I don't mean to suggest that this is the only interesting philosophical contribution she has made. There are a number of more specific claims that Korsgaard discusses which I find quite interesting, but since they are only tangentially related to my main focus, I will refrain from listing them here.

involves Korsgaard's view that the relationship between self-conceptions and practical reasons is unidirectional.<sup>2</sup>

There are actually two related problems here. First, this unidirectional account seems to get the grounds of at least some of our practical reasons wrong. According to Korsgaard, every time I have a normative reason to act, this reason will be grounded in an endorsed self-conception of mine. However, this claim is too strong given Korsgaard's understanding of what normative reasons are and where they come from. According to Korsgaard, normative reasons are the outcome of one's deliberations; they are a form of "reflective success." Based on this view of what a reason is (which I'll be discussing again in more detail in Section 5.4), it would seem as though determining *that* you have a normative reason would have to be closely related to seeing *why* you have this reason. This view of normative reasons, coupled with Korsgaard's claim that your normative reasons are grounded in endorsed self-conceptions, suggests that this ground of our reasons (namely our self-conceptions) should be quite apparent to us in our deliberations. However, this doesn't seem to be the case; I only occasionally take myself to have the reasons that I do because I hold certain self-conceptions. I may think, for example, that the reasons I have for saving a drowning stranger don't stem from any self-conceptions of mine.<sup>3</sup>

Second, a unidirectional view fails to account for an important aspect of this relationship. Frankfurt appears to be right in suggesting that endorsing a given course of action—and doing so without considering antecedent self-conceptions—can (and

---

<sup>2</sup> I have little doubt that Korsgaard thinks theoretical reasons can impact our self-conceptions—it is hard to see how she could deny this—but our interest involves the relationship between self-conceptions and practical reasons. While Korsgaard never explicitly denies that practical reasons can impact self-conceptions, at the very least she never says anything to suggest this (which, given her aim of developing a comprehensive account of this relationship, is itself a notable shortcoming).

<sup>3</sup> According to the view that I develop, we must reject both this account of what normative reasons are like and Korsgaard's claim that all reasons are grounded in self-conceptions. I argue against this view of reasons below, and go on to claim that while there is a close connection between normative reasons and certain self-conceptions (the presence of the former either is grounded in, or has important implications for, the latter), this relationship is not always a grounding relationship.

should, we may add) have some impact on the self-conceptions that one holds. I may, for example, decide that acknowledging that I have reason to help David get out of a jam requires that I start thinking of myself as “David’s Friend” rather than merely as “David’s Colleague.” A unidirectional account, however, fails to acknowledge this aspect of the relationship.

It is also worth noting that the central dilemma Korsgaard faces (discussed in Chapter 2) exists because of her commitment to the view that this relationship is unidirectional. This commitment (coupled with her insistence that the relationship holds universally) is precisely what leads to the “implausibility” horn of her dilemma. This diagnosis of Korsgaard’s dilemma, together with the two points above concerning what a unidirectional account gets wrong more generally, suggests that a bi-directional account of the relationship is a better alternative.

### **5.3 Frankfurt on the Relationship Between Self-Conceptions and Practical Reasons (Revisited)**

To review, Frankfurt’s account of the relationship in question is captured by CDT, which posits the following: an agent’s practical commitments and the normative reasons that are associated with them (on the one hand), and the agent’s self-conceptions (on the other), are mutually determining. While it is unclear precisely what sort of connection Frankfurt envisions between these items, he suggests that any determinations or discoveries that the agent makes with respect to one of them (normative reasons or self-conceptions) will have a significant impact on the other.

Since CDT posits a bi-directional relationship between self-conceptions and normative reasons, it can be split into two components. The first (reasons-to-conceptions) holds that in determining that I have reason to pursue object X, I will come to identify with my interest in X and will in turn wind up with a new or modified

self-conception. For example, if I come to think that I have reasons to help protect the environment (reasons that aren't derived from an antecedent self-conception), I thereby acquire the self-conception "Environmental Activist," or something to this effect.

The second component (conceptions-to-reasons) holds that when I conclude that I am simply the sort of person that cannot help but care about project Z (that this is part of who I am), I wind up with a new set of normative reasons that reflect this self-conception. For example, if I conclude that I am at bottom a "Pacifist," I will come to see that I have reasons for refraining from physically hurting others. Commitments that I make to a set of practical reasons become "encoded" in my self-conceptions, and vice versa; the two are in a sense different sides of the same coin.

How successful is this view? Beginning with what Frankfurt got right, his account of this relationship is an improvement over Korsgaard's simply insofar as it posits a bi-directional relationship. Not only is Frankfurt not committed to the view that all of one's normative reasons are grounded in endorsed self-conceptions, his view captures the fact that one's commitment to a practical reason can have an important impact on one's self-conceptions. While Korsgaard does a better job of both explicitly recognizing that there is a relationship and attempting to offer a systematic account of it, Frankfurt (who doesn't fare so well in these first two departments) manages to come closer to a correct account of this relationship (or at least a charitable reconstruction of his arguments comes close to such an account).

This brings us to the deficiencies in Frankfurt's account that were canvassed back in Chapter 3. While I will argue that the general spirit of Frankfurt's CDT is on target, I have already claimed that CDT is worrisome in four respects, three of which are relevant to the current discussion. The first and perhaps most serious concern was that CDT is under-specified. Frankfurt clearly thinks that there is a relationship

between normative reasons and self-conceptions, but he doesn't provide many details. Simply gesturing at a thesis of mutual determination isn't enough; we need to know more on how and why self-conceptions and normative reasons are mutually determining.

The second worry involved Frankfurt's suggestion that the conceptions-to-reasons component of CDT posits a link between self-conceptions and a very limited class of practical reasons: namely reasons of self-preservation. Frankfurt's argument for this component of CDT, it will be recalled, turns on the claim that I cannot act in ways that run contrary to what I take myself to be (or ideals that I set for myself) without jeopardizing my existence as the sort of thing that I am. If I come to the conclusion that I am at bottom a pacifist (which entails holding the self-conception "Pacifist"), then I have reason to refrain from performing acts of violence. Failing to acknowledge these reasons (i.e. failing, in Frankfurt's terms, to maintain a "certain motivational constancy or continuity" with respect to abstaining from violent behavior) will threaten my very survival, at least in a certain metaphorical sense, given that I take an essential feature of myself to be my pacifism. Frankfurt's claim is accordingly that my self-conceptions give rise to reasons primarily because a failure to act in certain ways might lead to my metaphorical death. It would seem, however, that self-conceptions can give rise to normative reasons that don't derive from worries about self-preservation, even of a metaphorical sort (as Korsgaard suggests).

The third worry was that Frankfurt's discussion of this relationship is couched in overly descriptive language. Frankfurt often appears to claim that there is a direct connection between reasons and self-conceptions, such that any change in one directly leads to a change in the other. Frankfurt remarks, for example:

When the decision is made without reservation....the person no longer holds himself apart from the desire to which he has committed himself....The

decision determines what the person really wants by making the desire on which he decides fully his own. To this extent the person, in making a decision by which he identifies with a desire, constitutes himself....He takes responsibility for the fact of having the desire...when he identifies himself with it” (Frankfurt (1): 170).

Frankfurt suggests that in making a decision—or in acknowledging that a certain reason for action exists—one *thereby* creates or modifies a certain self-conception (to put the claim in terms of self-conceptions, as I argued we should back in Chapter 3). As I will discuss in the next section, however, this largely descriptive account cannot cope with familiar counterexamples.

To this point we have established that Frankfurt’s account of the self-conceptions/normative reasons relationship is preferable to Korsgaard’s, at least in terms of its general structure. We have also seen, however, that Frankfurt’s account is not without its problems. In the next section, I will offer a first approximation of a better account of this relationship. Proceeding on the assumption that Frankfurt’s CDT is on the right track, I will begin by using CDT as a template, but will introduce modifications necessary to correct for the three specific problems discussed above.

#### **5.4 The Modified Co-Determination Thesis (MCDT): A First Approximation**

Since my own account of this relationship takes CDT as its starting point, I will hereafter refer to this modified view as MCDT to distinguish it from the view that I have attributed to Frankfurt. Note that in this section I will only be discussing a first approximation of MCDT; for reasons that will become clear shortly, we can think of this as a “subjective” version of MCDT, which I’ll refer to as MCDT<sub>S</sub>. In the following two sections, I will be making further changes to this first approximation to reflect the insights captured by Chapter 4 discussion of VBR and FBR.

I'll begin by laying out MCDT<sub>S</sub>, and will then turn to the task of justifying the move from CDT to this view. The central difference between MCDT<sub>S</sub> and CDT turns on the third objection discussed in the previous section. If MCDT<sub>S</sub> is to do a better job of modeling this relationship, it will have to present it in normative rather than descriptive terms. Our interest is in the norms that should govern the relationship between self-conceptions and normative reasons, not a claim that directly links one to the other as a matter of psychological fact or probability. For example, while we cannot simply say that by acknowledging a new reason for action one will *thereby* come to have a new self-conception profile (which is what CDT suggests), we can say that acknowledging a new reason for action implies that one *should*. MCDT<sub>S</sub> accordingly posits a connection between one's practical reasons and one's warrant for holding certain self-conceptions, at least when we are focusing on the reasons-to-conceptions component of this thesis.<sup>4</sup> Going in the other direction, MCDT<sub>S</sub> holds that one's acceptance of a self-conception implies that one *ought* to recognize certain considerations as normative reasons. According to this view, the relationship at issue is one of mutual implication; one's commitment to either a self-conception or a normative reason, in other words, has important implications for one's warrant in accepting the other. We can accordingly sum up MCDT<sub>S</sub> as follows:

- (A<sub>1</sub>) Conceptions-To-Reasons: Given that you hold a particular self-conception, there are certain associated considerations that you should treat as normative reasons.

---

<sup>4</sup> We needn't worry about warranted self-conceptions when going in the other direction—namely the conceptions-to-reasons direction—since instances of deliberation going this way will always involve an actual, as opposed to a “potential,” self-conception.

(B<sub>1</sub>) Reasons-To-Conceptions: Given that you treat certain considerations as normative reasons, you are warranted in holding associated self-conceptions.<sup>5</sup>

Two important notes. First, when discussing the A-leg of MCDT (both the subjective and objective versions that follow), I will be referring to general pro tanto reasons. For example, in the case of MCDT<sub>S</sub>, the claim is that holding the self-conception “Environmental Activist” implies that you ought to take yourself to have reason to recycle, etc. *generally* (even though you might not have reason to recycle something at this very moment). Second, since I will be making several modifications to these claims in what follows, I will use subscript numbers when describing MCDT’s components to keep distinct versions of this thesis separate.

To give some intuitive support for this view, let’s consider each component of MCDT<sub>S</sub> individually. First, it does indeed seem as though the fact that you hold a given self-conception counts in favor of thinking that you ought to act in certain ways. As I suggested in Chapter 1, part of what it means to hold many self-conceptions is for one to be committed to treating certain considerations as reason-giving. For example, one might wonder how much sense there could be in claiming that you hold the self-conception “Environmental Activist” if you thought that you didn’t have any reason to

---

<sup>5</sup> This is a good place to address a possible concern. One may worry that in some cases the implications of this second leg of MCDT will be rather uninteresting (this worry can actually arise with respect to either MCDT<sub>S</sub> or MCDT proper). For example, if I conclude that I have reason to eat chocolate, the self-conception that this reason warrants may simply be “Chocolate Lover.” One might wonder whether anything of interest follows from the claim that all reasons can be linked to self-conceptions in this way. Two points are worth making. First, trivial as these sorts of cases might appear, the implications at issue are nonetheless true. For you to think that you have a reason to perform a certain action, it does appear as though there must be some description of you that gives this reason purchase. Second, I’ll be discussing this general intuition about the link between self-conceptions and practical reasons in much greater detail in the next chapter. In particular, I’ll be arguing that MCDT models what turns out to be a conceptual connection between reasons and descriptions of possible respondents to them, and that this conceptual connection is what underwrites the demand that we be able to produce self-conceptions that link up to reasons in all cases (including fairly uninteresting ones, like that of our chocolate lover).

contribute time or money to environmental clean-up efforts or to recycle. As I'll discuss later, even self-conceptions that are not "inherently" action-guiding in this way (e.g. "5'10 White Male") can prove reason-giving in a more indirect sense (such that my holding the self-conception "5'10 White Male" counts against auditioning for part in a play that calls for a very tall Black actor).

The second component of  $MCDT_S$ , while perhaps less intuitive than the first, still has something to be said for it. This was the point of my earlier example in which seeing that I that have reason to help David get out of a jam may imply that I need to start thinking of myself as "David's Friend" rather than merely as "David's Colleague." Acknowledging that we have certain reasons for action appears to say something important about us and how we should think of ourselves.<sup>6</sup>

While a full discussion of the intuitions underwriting both components of  $MCDT$  will have to wait until the next chapter, let us focus for the moment on justifying the transition from  $CDT$  to  $MCDT_S$ . To do this, we need to show why the modifications to  $CDT$  that I've proposed are necessary if we are to correct for the difficulties that  $CDT$  runs into. Let us start by returning once more to the third objection of the previous section, which concerned the overly descriptive character of  $CDT$ . The problem here is that one can easily think of cases in which an agent might act in pursuit of a practical commitment without it having any real impact on her self-conceptions. One might not reflect at all on the impact that a set of decisions might have on one's self-conceptions, or alternatively, one might commit oneself to a line of conduct that carries implications for one's self-conception that one finds too troubling to accept. For example, what if an agent stubbornly refuses to think of himself as a "Selfish Liar" despite the fact that he lies all the time for personal gain? (Perhaps he

---

<sup>6</sup> While this claim might seem a bit vague at this point, keep in mind that several sections of Chapter 6 will be dedicated to explaining and justifying it.

can't bring himself to think of himself that way, but is greedy enough to remain committed to a life of lies). What if one's actual self-conceptions don't track one's practical reasons in a coherent way? Not only is it the case that some act of reflectivity on the agent's part might be necessary for one's practical commitments to impact one's self-conceptions, but furthermore, these acts of reflectivity can at times "malfunction," in that the self-conception that the agent forms might not coherently reflect her practical commitments. CDT is simply not set up to handle these cases.

MCDT<sub>S</sub>, on the other hand, has no problem with them. Take the example of a person who lies all the time for his personal gain but holds the self-conception "Selfless Purveyor of the Truth." MCDT<sub>S</sub> suggests that this agent is not warranted in holding this self-conception, but is rather warranted in holding the self-conception "Selfish Liar," or something to this effect (given our understanding of the self-conceptions in questions). Conversely, given that an agent is committed to the self-conception "Susan's Friend," MCDT<sub>S</sub> suggests that this agent *should* treat certain considerations—for example, that Susan is in distress—as counting in favor of certain actions. Susan's friend ought to treat these considerations as reason-giving because that's part of what it means to be someone's friend and to see oneself as such.

How does MCDT<sub>S</sub> fare against the other two objections that were raised against CDT in the previous section? With respect to the "lack of detail" worry, even on the basis of the minimal discussion offered to this point, MCDT<sub>S</sub> seems to fare somewhat better. Rather than simply gesturing at a vague relationship between self-conceptions and reasons for action, it posits a more detailed bi-directional relationship between the two by showing how one's commitment to one item (reasons or self-conceptions) carries important implications for the other. However, should one still be worried that MCDT<sub>S</sub> doesn't say enough about the mechanics of this relationship (and in particular, about the epistemic standards underlying the warrant at issue), we can

certainly supplement this theory with additional MCDT-friendly models to provide the requisite detail.

To give an example of how one might do this, recall the notion of an “extension set” which was introduced in Chapter 1 as a way of formalizing the self-conceptions/normative reasons relationship. This notion has its origin in the observation that many self-conceptions are closely associated with characteristic activities and concerns. The self-conception “Environmental Activist,” for example, is associated with the performance of environmental clean-up efforts, ensuring that one purchases only recyclable materials, etc. Self-conceptions of this sort are characterized, and even constituted in large part, by activities and concerns of this general kind. To understand what an environmental activist is requires an understanding of the sort of practical principles that are acknowledged by people falling under this description. I have defined the set of practical principles that are associated with a given self-conception as that self-conception’s extension set.

Linking this claim with  $MCDT_S$ , the implication is that one’s warrant for holding a given self-conception will be determined to some extent (and at a bare minimum) by the congruence between the facts that obtain in a particular individual’s case (to include facts about the individual’s commitment to certain ends) and the extension set of the given self-conception. If one’s practical commitments are at odds with the members of a self-conception’s extension set, this may be sufficient to rule out one’s warrant for holding this self-conception.<sup>7</sup> While this sketch may be too brief

---

<sup>7</sup> It should also be noted that one’s warrant for holding a given self-conception may be determined in part by more objective facts as well, to include the skill and abilities of the agent, her legal standing, and the subjective commitments of others (e.g. perhaps the truth of the self-conception “Friend of Susan” requires not only certain commitments on my part, but certain commitments on Susan’s part as well). To sum up, we can say that in general, an agent A is warranted in holding self-conception S if (a) A’s practical commitments are consistent with S’s extension set in the right way, and (b) A and her circumstances satisfy whatever objective requirements (if any) attach to S.

to provide the sort of detail one might ask for, one could certainly add more detail to models of this sort to bolster the explanatory power of  $MCDT_S$  as needed.

This brings us to the final worry that CDT posits an overly narrow account of the way in which self-conceptions can ground practical reasons (namely, by implying that all such reasons are grounded in the value of self-preservation). Recall that on Frankfurt's view, I have reason to act so as not to falsify an endorsed self-conception because in so doing, I avoid facing a certain sort of metaphorical "death" as a person rightly described under that description. However, even on a very intuitive level, this isn't the sort of justification that we tend to offer for not acting against our self-conceptions. As stated,  $MCDT_S$  (which we needn't wed to a preoccupation with metaphorical self-preservation) leaves open the possibility of alternative explanations of the conceptions-to-reasons component of the thesis. That my commitment to the self-conception "Justin's Brother" can provide me with reasons for acting in certain ways toward Justin is itself intuitive absent any reference to claims that I might "perish" in some sense as his brother if I fail to do so. As suggested above, part of what it means to accept the self-conception in question is to acknowledge certain considerations as reasons for action. (Put another way, part of what it means to accept a self-conception is to accept the principles contained in its extension set, which in turn demands that one acknowledge certain reasons for acting.)

Since I have suggested that the bi-directionality of Frankfurt's account is an improvement over the uni-directionality of Korsgaard's account, it should come as no surprise that the bi-directional MCDT represents an improvement over her account as well.  $MCDT_S$  agrees with Korsgaard in claiming that one's commitment to a given self-conception will give rise to certain reasons for action (thus establishing a conceptions-to-reasons connection).  $MCDT_S$  goes on to add, however, that a given agent can nonetheless identify reasons for acting that don't stem from any

consideration of her self-conceptions, reasons that moreover carry strong implications for the agent's warrant in holding certain self-conceptions (thus establishing a reasons-to-conceptions connection). Responding directly to the dilemma Korsgaard faced, MCDT<sub>S</sub> allows us to claim that *every time* an agent concludes that she has a reason to act, we will be able to say something interesting about the relationship between this reason and her self-conceptions (since *either* the reason will be grounded in an antecedent self-conception, *or* the reason will have important implications for her warrant in holding certain self-conceptions). Thus, in the face of Korsgaard's dilemma, one might be justified in claiming that there is a plausible and non-trivial relationship between one's practical reasons and one's self-conceptions.

### 5.5 VBR and FBR Revisited

With this first approximation of MCDT in hand, let us now turn to the question of how the Value-Based Reasons thesis (VBR) and the Fact-Based Reasons Thesis (FBR) impact the views we have been examining. Having made the right "structural" modifications to CDT already (by clearing up the bi-directionality issue, ensuring that the relationship posited is one of justification rather than psychological tendencies, etc.), we have to make sure that we have properly specified the scope of this relationship. For as it turns out, this relationship doesn't hold between normative reasons and just any old self-conceptions. Before turning to this claim, however, I will first discuss how Korsgaard's and Frankfurt's accounts run afoul of VBR and FBR. Doing so may illuminate other potential problems that MCDT must avoid. In the next section, I will apply these two theses to MCDT<sub>S</sub> to develop a more accurate account of this relationship.

As a quick review, recall that VBR maintains the following: when a normative reason for me to  $\phi$  exists, it does so at least in part because a given desire-independent

value exists. In other words, it is a necessary condition on R's being a normative reason for me to  $\phi$  that  $\phi$ -ing possess value-making or value-constituting properties that obtain independent of whether I or a more rational version of myself care about them. This conclusion was justified in part through a survey of our phenomenological commitments. This survey revealed that upon reflection, we find it hard to justify a course of action to ourselves or to others (i.e. we have a hard time believing that a normative reason to perform a certain action exists) unless we can link this action to desire-independent values. (Also important to this justification was an argument that we must take these phenomenological considerations seriously).

Similar phenomenological considerations were taken to underwrite FBR, which holds that our reasons are based in facts about our circumstances, rather than our mere beliefs about these circumstances. While we inevitably act on the basis of our beliefs—such that every time we act on a normative reason we do so based on our beliefs about our situation—from our own perspectives as deliberating agents, it is not our beliefs as such that underwrite these reasons, but rather facts (to include facts about the self-conceptions one holds). Support for FBR comes largely by looking at cases in which we have mistaken beliefs; while we might cite these mistaken beliefs to excuse poor actions on our part, what ultimately counts in favor of certain actions (in our own eyes) are underlying facts.

### **5.5.1 Implications for Korsgaard's Account**

How do these two theses bear on Korsgaard's account? The first problem her account runs into—which doesn't directly involve VBR, but rather the general picture of normative reasons that VBR recommends—concerns her view of normative reasons for action. According to Korsgaard, normative reasons are simply the outcome of an agent's practical deliberations. She writes: "We need reasons because our impulses

must be able to withstand reflective scrutiny. We have reasons if they do. The normative word ‘reason’ refers to a kind of reflective success....[Claims that one has reason to act] mean the work of reflection is done” (Korsgaard (1): 93-4). Korsgaard thus paints a “projectivist” view of normative reasons rather than the “detectivist” view that VBR recommends.

According to Korsgaard, it doesn’t make sense to speak of an agent having a normative reason prior to the agent’s determination that she has such a reason. However, on the basis of the discussion in the preceding chapter, this is not the case. From my own point of view as a deliberating agent, my aim is to figure out which considerations speak in favor of acting in certain ways, and from this deliberative point of view, it will often be the case that certain considerations speak in favor of performing certain actions prior to my awareness of them. For example, we are inclined to think that an agent can have a normative reason to get an education or act in ways that preserve one’s long-term health regardless of whether he himself has come to this conclusion. But on Korsgaard’s account of normative reasons, this isn’t so.

The upshot is that we must reject what I’ve called Korsgaard’s expression claim (“My normative reasons are expressive of my self-conceptions”) and the implications that it carries in the steps that follow (to include the Step Three interpretation of this claim). The reason is that this claim will only be true in the case of motivating reasons (and even then, only in the trivial sense in which a reason expresses what I referred to in Chapter 2 as a “thin” self-conception). It is true that any bit of practical reasoning that I perform will be expressive of me in at least a weak sense—and so, accordingly, will be the motivating reasons that stem from my deliberations—but many of the *normative* reasons that I am trying to get at are not directly self-expressive, since many of them exist apart from any deliberative act on

my part. According to the VBR account of reasons, I can have a normative reason to help someone in distress without being aware that this reason exists. If this is so, in what sense can this reason possibly be “expressive” of me?

A second problem for Korsgaard’s account involves, unsurprisingly, her commitment to DBR. Recall that the general import of VBR was that normative reasons for action aren’t grounded directly in our desires. While desires sometimes do play a role in specifying which of a given set of candidate objects one has an all-things-considered normative reason to pursue, (a) the initial candidacy of these options will be determined by desire-independent values, and (b) the desire can “tip the scales” in favor of one option only by way of other desire-independent values (like satisfaction) that the desire puts one in touch with.

Based on Chapter 2 discussion, it seems fairly clear that Korsgaard subscribes to DBR, but perhaps it would help to offer more support for this claim. Recall that DBR maintains the following: every normative reason for action that exists does so (a) because a certain desire exists, and (b) not because a certain desire-independent value exists. Beginning with the first component of this claim, Korsgaard clearly holds that desiring a certain outcome is a necessary condition for one’s having a normative reason to pursue that outcome; this is, after all, what her Step Four claim maintains.

With respect to the second component of DBR, in a Chapter 4 footnote it was pointed out that subscription to this part of DBR is consistent with the following two positions: (1) the view that there are no desire-independent values, and (2) the view that there are desire-independent values, but they do not ground normative reasons. Does Korsgaard subscribe to one of these two views? On the one hand, Korsgaard’s subscription to a constructivist theory of value suggests that (1) may be her considered view, insofar as her constructivism implies that there are no desire-independent values. She claims at one point that: “Good maxims are intrinsically normative entities, but

they are also the products of our own legislative wills. In that sense, values are created by human beings...the maxim isn't a law until we will it, and in that sense create the resulting values" (Korsgaard (1): 112).

On the other hand, Korsgaard's (notably ambiguous) discussion of the value of humanity leaves open the door for the argument that she subscribes to (2), in that the value of humanity might be thought of as a desire-independent value, but one that has no direct bearing on one's *practical* reasons. (Her considered view, once again, appears to be that we ought to endorse our moral identity as a practical identity, but that we have no normative reasons to act in accordance with our moral identity until we have endorsed it). In either event, it is clear that Korsgaard is committed to one of these two possibilities, and accordingly, DBR.

The upshot is that VBR undermines Step Four of Korsgaard's argument ("My normative reasons are grounded in objects of my concern (things that I desire or care about)"). From my own perspective as a deliberating agent, not only is it the case that the mere fact that I care about a given self-conception fails to make it reason-giving in my own eyes; as it turns out, nor does the mere fact that I *don't* care about a given self-conception rule out the possibility that it is reason-giving for me. This second point indicates another respect in which Korsgaard's reliance on DBR yields the wrong results. As a consequence of her commitment to DBR, Korsgaard incorrectly limits her discussion of the relationship between practical reasons and self-conceptions to a small subset of the latter, namely those that one cares about (i.e. practical identities).

To illustrate this problem, that I take myself to be an "Insensitive, Short-Tempered Friend"—that I think that this description is true of me—appears to give me a normative reason to avoid being around my friends when they are undergoing what is to them a great hardship, but to me a trifling worry of no consequence. Given that I

am this way, I take myself to have reason to steer clear of friends in these situations because of the likelihood that I may say something hurtful. This self-conception accordingly appears to be practically relevant even though I don't endorse it. On the basis of these points, we must also reject Korsgaard's Step Six claim ("My normative reasons are grounded in my practical identities"), the validity of which depends on Step Four.

### 5.5.2 Implications for Frankfurt's Account

We turn now to Frankfurt's account of normative reasons, and in particular, those elements of it that conflict with VBR. The first point of concern involves the incompatibility of VBR and the first-level component of Frankfurt's account (which relies on what I referred to Frankfurt's Internalism Thesis). It should now be clear that this first-level account runs afoul of VBR by suggesting that from the first person, deliberative point of view, an agent will take the existence conditions of his normative reasons to depend on his particular desires in every case, and not on the desire-independent values in play (if there are such things). According to Frankfurt, normative reasons for me exist only because either I hold a self-conception that I care about, or I am committed to a ground project that I care about. According to VBR, however, this is not so. From the point of view of the deliberating agent, one's reasons may depend indirectly upon one's desires (but not in all cases, and never *qua* desire-satisfaction), but at the very least they require desire-independent values. The upshot is that if we accept the arguments presented in the preceding chapter, we must reject the general conclusion of the first-level component of Frankfurt's account of normative reasons.

As it turns out, we have good reason to reject Frankfurt's second-level account as well. First, Frankfurt hasn't given us sufficient reason to think that

wholeheartedness is the only object of intrinsic, final value. His argument rests on (a) the plausibility of claiming that wholeheartedness is a reason-giving value for everyone, together with (b) a general skepticism that anything else is universally reason-giving in this way. If, however, there is enough overlap in our views of what has desire-independent value (which is what Chapter 4 argued), then we have good reason to reject this argument. Being healthy, for example, seems to be a desire-independent value that is pro tanto reason-giving. If this is so, then wholeheartedness is not the only object of desire-independent value, and Frankfurt's second-level argument is in trouble. What if my health and wholeheartedness conflict?

Where does FBR fit into all of this? Neither Korsgaard nor Frankfurt explicitly claim that only true factive self-conceptions can be reason-giving, but clearly it must matter whether or not a given self-conception is based on underlying facts. Given what we've been told, neither thinker has explicitly ruled out the possibility that false factive self-conceptions can be reason-giving, so long as they are endorsed by the agent. Clearly, however, we don't want to say that an out of shape 45 year old with the self-conception "World Class Athlete" has normative reason to run in the NYC marathon, regardless of how important this description is to him. Given how obvious this point is, I am inclined to think that Korsgaard and Frankfurt would want to rule out these sorts of cases. It appears, however, either that it didn't occur to Korsgaard or Frankfurt to comment on these sorts of cases, or they thought it was simply obvious that false self-conceptions can't be reason-giving. Given the importance, prevalence, and complexity of these sorts of cases, however, it would be a good idea to state how one's account of the relationship deals with them.<sup>8</sup>

---

<sup>8</sup> I am inclined to stop short of saying that Korsgaard and Frankfurt do (or reflectively would) subscribe to FBR, despite the fact that they would most likely want to rule out cases like the out of shape marathon runner. While they might rely on something like FBR to rule out this sort of case, they could also do so by insisting (for example) that this out of shape man has other, stronger desires that suffering from a heart attack would frustrate.

## 5.6 Applying VBR and FBR to MCDT: A Second Approximation

Our task now is to apply these two theses to our first, subjective approximation of MCDT—MCDT<sub>S</sub>—which holds the following:

(A<sub>1</sub>) Conceptions-To-Reasons: Given that you hold a particular self-conception, there are certain associated considerations that you should treat as normative reasons.

(B<sub>1</sub>) Reasons-To-Conceptions: Given that you treat certain considerations as normative reasons, you are warranted in holding associated self-conceptions.

Let us begin by looking back at the Chapter 4 approach to making sense of normative reasons for action. This approach, it will be recalled, involved thinking about normative reasons from the first-person, deliberative point of view. On the assumption that this approach is legitimate (it was justified in the second half of the previous chapter, and will receive further support in the next chapter), the first modification we need to make to MCDT<sub>S</sub> is to reframe the antecedent of (B<sub>1</sub>) to properly reflect our first-person, deliberative commitments. Reflection reveals that our deliberative interest isn't in the implications of what we simply *think* are normative reasons, but rather in the implications of our *actual* normative reasons. From my point of view as a deliberating agent, I don't care about the implications of false or non-existent reasons on my warrant for holding certain self-conceptions; if what I think are normative reasons actually aren't, I'm not interesting in their implications at all. I'm trying to get things right, and so my interest is only in actual normative reasons. We accordingly need to modify (B<sub>1</sub>) to reflect this fact.

In the interest of staying consistent to this “first-person deliberative interest” theme, we can also modify the consequent of (A<sub>1</sub>) to reflect the fact that our

interest isn't just in what we should treat as normative reasons, but more simply in what normative reasons we in fact have. Note that this shift in treatment away from a purely "subjective" reading also forces us to replace talk of "warrant" with talk of "rendering appropriate" (we are now concerned with objectively permissible inferences, not inferences that one might be justified in drawing based on whatever evidence happens to be available to one). This first set of modifications gives us the following:

- (A<sub>2</sub>) Conceptions-To-Reasons: Holding a certain self-conception gives you associated normative reasons for acting in certain ways.<sup>9</sup>
- (B<sub>2</sub>) Reasons-To-Conceptions: Having certain normative reasons for acting makes it appropriate for you to hold associated self-conceptions.

The next set of modifications involves a more direct application of VBR and FBR. At this point, however, things begin to get a bit complicated. Recall that the general term "self-conceptions" refers to two distinct classes of descriptions: factive self-conceptions (which are descriptions that an agent thinks are true of himself) and ideal self-conceptions (which are descriptions that an agent thinks are false, but is working to make true). When inquiring into the relationship between normative reasons and these categories of self-conceptions, we need to examine a number of different possibilities. In particular, there are five specific categories of self-conceptions that we need to look at:

- (a) True, Value-Commensurate<sup>10</sup> Factive Self-Conceptions.
- (b) True, Value-Incommensurate Factive Self-Conceptions.
- (c) False Factive Self-Conceptions (regardless of their relation to values).
- (d) Value-Commensurate Ideal Self-Conceptions.

---

<sup>9</sup> Once again, this should be understood as applying to general pro tanto reasons, and not necessarily all-things-considered reasons that apply in very specific circumstances.

<sup>10</sup> Again note that the term "value" in this use refers to objective values, not a particular agent's "values."

## (e) Value-Incommensurate Ideal Self-Conceptions.

The question to ask at this point is: what do VBR and FBR tell us about how our normative reasons are related to these different sorts of self-conceptions?

Let us begin with ideal self-conceptions. Based on our discussion of VBR, it should be fairly clear that value-commensurate ideal self-conceptions give one pro tanto normative reasons for acting in certain ways. On the assumption that the ideal self-conception “Caring and Understanding Friend” directs one to act in value-commensurate ways (perhaps the agent in question is fairly insensitive at the moment), one might have good reason to work to become a “Caring and Understanding Friend,” and to treat one’s friends appropriately.

Alternatively, VBR just as clearly holds that value-incommensurate ideal self-conceptions are not reason-giving (either in the sense of recommending practical projects—except perhaps a project of getting rid of this self-conception—or in the sense of being practically relevant). Imagine that I hold the ideal self-conception “The Great Hope of Oppressed Minority X.” As it turns out, however, not only am I an unlikely candidate for becoming a Great Hope (perhaps I am too inept to champion any great cause and it is unlikely that anyone would ever look up to me as a Great Hope), perhaps this ideal encourages the pursuit of countervalueable activities. Perhaps, for example, the goals embodied by this particular ideal require acting in profoundly disrespectful ways toward the people who will be the target of my efforts.

The important point is this. Even if I care tremendously about this Great Hope project, VBR tells us that it cannot be reason-giving. This is contrary to the general view espoused by Korsgaard and Frankfurt. Based on what they say, it would seem that I have good reason to go around interfering in the lives of Minority X members, simply because this ideal means so much to me. As we have seen, however, from my own perspective as a deliberating agent, I don’t want to engage in activities that are in

fact countervailing. It might be tough for me to come to grips with the fact that I need to abandon this project that means so much to me, but it will appear to me as something that I simply have to do (at least considering matters in a cool hour). We must accordingly modify MCDT as follows:

(A<sub>3</sub>) Conceptions-To-Reasons: Holding a value-commensurate ideal self-conception gives you associated normative reasons for acting in certain ways.

(B<sub>3</sub>) Reasons-To-Conceptions: Having certain normative reasons for acting makes it appropriate for you to hold associated value-commensurate ideal self-conceptions.

This brings us to the three possible factive self-conceptions listed above (a–c). Beginning with the third, FBR suggests that false factive self-conceptions are not reason-giving.<sup>11</sup> The idea behind this requirement is fairly intuitive. To return to an earlier example, we don't want to say that an overweight, middle-aged man in poor health has normative reason to run in the NYC marathon in light of the false factive self-conception "World Class Athlete." Generally speaking, a plan to perform an action that is formed in light of false claims about what the world is like is bound to fail. The intuitive conclusion is that any claim of the form "I have normative reason to  $\phi$  *because* I am an X" will be false if I am not in fact an X.<sup>12</sup>

---

<sup>11</sup> It is worth noting that a shift in one's cognitive attitude toward a false, value-commensurate factive self-conception will turn it into a value-commensurate ideal self-conception (i.e. the only difference between the two involves one's beliefs about whether the description in question is true or ought to be made true).

<sup>12</sup> One might think that the falsity of a given factive self-conception can nonetheless be practically relevant. For example, perhaps knowing that the self-conception "World Class Athlete" is false suggests that I shouldn't run in marathons. It might be useful, however, to distinguish between the claim that false factive self-conceptions can be practically relevant and the claim that the contrapositive of such self-conceptions can be. In this case, I'm inclined to think that it is the corresponding true factive self-conception "Out of Shape Middle Aged Man" is what proves practically relevant in this case.

One possible complication concerns cases in which acting on a false self-conception might actually lead to positive outcomes. While a more pragmatic approach to this subject would suggest that one has normative reason to act on a false self-conception if it will lead to better results, this approach is problematic given our first-person, deliberative approach to these issues. It would be very difficult for me to propose (to myself) acting on the basis of false factive self-conceptions. In any given case, either I will know that a given factive self-conception is false but value-commensurate—in which case I should properly treat it as a value-commensurate ideal self-conception—or I will need to deceive myself into thinking that a false factive self-conception is true (perhaps because, even though I know it isn't generally value-commensurate, I see some benefits to holding it in a particular case). The first alternative is consistent with denying false factive self-conceptions a reason-giving role, and the second alternative raises the problematic issue of rationally deceiving oneself. While acknowledging that more could be said on this point, I will proceed on the assumption that we can dispense with recommendations that one deceive oneself in these ways (especially since the sort of cases at issue—i.e. ones in which it is beneficial to hold a false factive self-conception that isn't generally value-commensurate—are probably quite rare).

So much for false factive self-conceptions. What about true factive self-conceptions? As it turns out, both sorts of true self-conceptions—those that are value-commensurate and those that aren't—are related to one's practical reasons in interesting ways. Focusing on the conceptions-to-reasons element of this relationship, this will be so because true factive self-conceptions can contribute to one's deliberations in two different ways. First, true value-commensurate factive self-conceptions can recommend practical projects. For example, given that I hold the true self-conception "Community Activist", and since this is a valuable thing to be (it

accords with desire-independent values, let us assume), this self-conceptions will recommend appropriate practical projects.

Second, true self-conceptions can prove practically relevant by, for example, figuring into one's assessments of the likelihood of succeeding in certain endeavors. The fact that the true, value-incommensurate self-conception "Out of Shape Couch Potato" is true of me tells me that I ought to call for help rather than attempt to chase down a purse-snatcher, even though I don't value my life under this description and in fact wish that it was false. Disliking this true self-conception might lead me to acquire the ideal self-conception "Fit Athlete," but until this ideal is realized (and so the true self-conception "Out of Shape Couch Potato" is rendered false), this true description of me will remain practically relevant in the requisite sense. Given these two roles, we can claim more generally that true factive self-conceptions, regardless of their relation to values, play a reason-giving role, although in two different ways.

To head off a possible worry, it is worth noting that the factors that determine the "truth" of a self-conception are quite varied. Two points that were discussed in previous chapters are worth recalling. First, the set of true self-conceptions that one can hold at any given time is underdetermined by facts about what one is. This is so because one can often describe a set of underlying facts in a number of different ways (by thinking of different "groupings" of these facts as representing different self-conceptions). The upshot is that there is no absolute one-to-one correlation between underlying facts (or groups of facts) and what can count as a "true" self-conception.

Second, recall that some self-conceptions will be made true simply in virtue of an agent's attitudes and interests. That I really enjoy following soccer may be enough to make true the self-conception "Soccer Enthusiast." Furthermore, while not all self-conceptions fall into this group, most have truth conditions that are satisfied at least in part by the choices of the agent. The self-conception "Susan's Friend," for example,

may be made true in part by my attitudes and actions toward Susan, and in part by her attitudes and actions toward me (on the assumption that true friendship requires reciprocal feelings, etc.). These facts should assuage fears about the austerity of the claim that normative reasons are tied to factive self-conceptions only if they are true, since this claim doesn't imply that your reasons are limited to facts about your situation in life that you have no control over.

Updating MCDT to reflect these claims about factive self-conceptions, we get the following:

(A<sub>4</sub>) Conceptions-To-Reasons: Holding a value-commensurate ideal self-conception or a true factive self-conception gives you associated normative reasons for acting in certain ways.

(B<sub>4</sub>) Reasons-To-Conceptions: Having certain normative reasons for acting makes it appropriate for you to hold associated value-commensurate ideal self-conceptions or true factive self-conceptions.<sup>13, 14</sup>

---

<sup>13</sup> To put this point in more general terms (i.e. dropping any mention of factive vs. ideal self-conceptions), MCDT holds the following: self-conceptions that are either true OR value-commensurate stand in an interesting relationship to normative reasons, whereas self-conceptions that are both false AND value-incommensurate do not (the only reasons they give are reasons to stop holding them). Beginning with the former, there are three possible categories of licensed self-conceptions that stand in an interesting relationship to normative reasons. (1) True, Value-Commensurate Self-Conceptions. For example, if I rightfully hold the self-conception "Environmental Activist," and if this self-conception is value-commensurate, it seems that this self-conception will be related to my normative reasons in interesting ways. (2) True, Value-Incommensurate Self-Conceptions. Based on the distinction drawn earlier, we can also see that self-conceptions of this sort can affect one's normative reasons as practically relevant considerations. If it turns out that I hold the true self-conception "Serial Killer," while this fact doesn't give me normative reason to kill people (since it is value-incommensurate), it may serve as a practically relevant consideration by giving me reason to avoid a situation in which I might be tempted to kill someone. (3) False But Value-Commensurate Self-Conceptions. Finally, it also turns out that a self-conception of this sort can rightfully figure into my practical ends despite the fact that it is false. For example, I might be working to make true the self-conception "Community Activist." While this self-conception is not yet true of me, holding it may give me reason to start looking for ways to become more involved in the community and to learn more about good leadership so that I can make this ideal self-conception true. On the other hand, a self-conception that is both false and value-incommensurate cannot play either of these roles. For example, if the self-conception "Terrorist" is not yet true of me, and if it is characteristically associated with counter-valuable activities (like killing innocent people), this self-conception can neither figure into any of my reasons for acting (except perhaps a reason to stop holding this self-conception), nor can any of the normative reasons that do apply to me render it appropriate for me to hold this self-conception.

In the interest of making these rather cumbersome claims a bit more manageable, let us refer to any value-commensurate ideal self-conception or true factive self-conception as a “licensed self-conception” (because—for reasons that will become clear shortly—agents will be licensed in holding them and making them operative in their practical reasoning). Substituting this term into our thesis, we get the following:

(A<sub>5</sub>) Conceptions-To-Reasons: Holding a licensed self-conception gives you associated normative reasons for acting in certain ways.

(B<sub>5</sub>) Reasons-To-Conceptions: Having certain normative reasons for acting makes it appropriate for you to hold associated licensed self-conceptions.

While we are now very close to a final version of MCDT, there is one more issue we must address. This version of MCDT implies that every normative reason I have that is associated with a licensed self-conception exists *because* I hold the licensed self-conception in question. As we have seen, however, normative reasons for acting can exist for an agent even if she doesn’t hold a licensed self-conception that is associated with them (our analysis of Korsgaard made this clear). I cannot, for example, dodge the normative force of an obligation to take care of my son by simply

---

<sup>14</sup> As this claim might seem a bit odd, it is worth discussing the sense in which the existence of a normative reason makes it appropriate to hold a true factive self-conception. To give an example that I’ll return to in upcoming discussion, generally speaking the fact that I have normative reason to help John paint his house this week may “make it appropriate” for me to hold the self-conception “John’s Friend,” in the sense that this reason couldn’t apply to me if I wasn’t John’s Friend (its existence entails the appropriateness of holding this self-conception). However, how best to describe the relationship between (a) underlying facts about a case, (b) the normative reasons associated with these facts, and (c) the self-conceptions made appropriate by (a) and/or (b) is no easy task. Are the underlying facts prior to the normative reasons (in a way that would make it misleading to the claim that the normative reason *itself* makes the self-conception appropriate), or does the nature of the supervenience relationship between the two make this claim acceptable? According to the former view, we could say that the facts about me that make it the case that I have normative reason to help John are also what make it the case that it is appropriate for me to hold this self-conception, and so the normative reason is acting as a “middle man” in this case. (However, it is worth pointing out that even if this view is correct, in many cases the normative reason may play a rather significant epistemic role by serving as the consideration that first signals to the agent that a certain self-conception would be appropriate to hold). Regardless of which understanding of the facts/normative reasons relationship one favors, the claim made in B4 is nonetheless true; it is simply worth noting that there may be some question as to whether it is the normative reason *itself* that renders the factive self-conception appropriate.

refusing to hold the self-conception “Nicholas’ Father;” my reason for taking care of Nicholas is independent of my self-conceptions. But if this is so, then contrary to what (A<sub>5</sub>) implies, holding the licensed self-conception “Nicholas’ Father” can’t be what *gives* me this normative reason.

To work out point, it might help to distinguish between two sorts of cases. First, there are cases in which underlying facts about me give me normative reasons that are independent of my self-conceptions. The Nicholas’ father example is one such a case. To give another example, given that I am a rational being, I might have reason to act in accordance with the Categorical Imperative, even if I don’t hold the self-conception “Rational Being.” This is a case in which my self-conceptions don’t necessarily figure into the existence of my normative reasons, at least on a certain understanding of Kant’s project. (Note, however, that based on MCDT, this doesn’t mean that we can’t say anything interesting about self-conceptions in such cases. The second leg of MCDT tells us that the existence of this reason makes it appropriate for me to hold the self-conception “Rational Being.”)

Call these “antecedent reason cases.” They occur whenever underlying facts give rise to normative reasons independent of whether one holds an associated licensed self-conception. When this happens, the self-conceptions that are associated with these independent reasons—if held by the agent in the first place—don’t *give rise* to these normative reasons, but rather *indicate* their pre-existence. In these cases, the underlying facts don’t leave open a menu of different licensed self-conceptions (each of which may be associated with different normative reasons) for me to choose between.

Alternatively, there are cases in which the underlying facts in question do underdetermine the self-conceptions that it would be appropriate for me to hold (and hence the normative reasons that I have). For example, suppose that I work with John

and the underlying facts in question leave it open to me to think of myself merely as “John’s Colleague” or as “John’s Friend.” In this example, perhaps the fact that I think of myself as “John’s Friend” *makes it the case* not only that this is a true description of me, but also that I have certain normative reasons for acting (like a reason to help John paint his house this weekend). In these “posterior reason cases,” it would be appropriate to say that holding a licensed self-conception is what gives rise to normative reasons for acting.

Updating our model to reflect this difference in cases, we get the following:

(A<sub>6</sub>) Conceptions-To-Reasons: Holding a licensed self-conception either gives you associated normative reasons for acting in certain ways (in posterior reason cases) or indicates the pre-existence of associated normative reasons for acting in certain ways (in antecedent reason cases).

(B<sub>6</sub>) Reasons-To-Conceptions: Having certain normative reasons for acting makes it appropriate for you to hold associated licensed self-conceptions.

To this final substantive modification we can add a final cosmetic modification that generalizes the two cases mentioned in the rather unwieldy (A<sub>6</sub>) claim. The result is the following, final version of MCDT:

**The Modified Co-Determination Thesis (MCDT):**

(A) Conceptions-To-Reasons: Holding a licensed self-conception entails the existence of associated normative reasons for acting in certain ways.

(B) Reasons-To-Conceptions: Having certain normative reasons for acting makes it appropriate for you to hold associated licensed self-conceptions.

This restatement of the conceptions-to-reasons element of MCDT tells us more generally that when an agent holds a licensed self-conception, he will have normative reasons for acting that are associated with this self-conception. This way of putting the point covers cases in which the licensed self-conception gives rise to the associated

normative reason, and cases in which it simply indicates facts that independently give rise to the associated normative reasons.

With this final version of MCDT in hand, we can see that this model is basically an “implications function” that takes as an input the member(s) of one class of items—either a commitment to a licensed self-conception or a fact about the normative reasons that one has—and provides as an output this input’s implications for the member(s) of the other class of items.

To update a point that was made earlier, one upshot of MCDT is that every time an agent acts on the basis of a normative reason, we can say something interesting about his self-conceptions. Either his action will have been prompted by consideration of his licensed self-conceptions, or his acting for this reason will have important implications for the appropriateness of holding certain licensed self-conceptions.

### **5.7 Practical Principles Governing Self-Conceptions and Their Application**

On the basis of this discussion, we can now lay out some general practical principles that have been emerging from our discussion. These principles tell us which considerations are relevant to our questions about (1) how our self-conceptions impact our normative reasons and (2) what sorts self-conceptions we ought to hold. They accordingly govern the maintenance and “operative status” of self-conceptions when it comes to answering these sorts of questions. The practical principles in question are as follows:

1. Acquire true factive self-conceptions and discard false factive self-conceptions.

2. Make all true factive self-conceptions operative in your practical reasoning. In this application, factive self-conceptions contribute to the deliberative process by either (a) setting ends for us (in the case of factive self-conceptions that are value-commensurate), or (b) being practically relevant to our assessments of what we ought to do.<sup>15</sup>
3. For all true factive self-conceptions that are value-incommensurate, work to make them not true and then discard them per (1), but per (2b), continue to make them operative in your practical reasoning for as long as they are true.
4. Acquire only ideal self-conceptions that are value-commensurate and discard ones that are not.
5. Make accepted, value-commensurate ideal self-conceptions operative in your practical reasoning.

With this list of principles in hand, a few explanations and qualifications are needed. First, by “operative in one’s practical reasoning,” I mean that the agent should take note of the self-conception when deliberating about what to do, at least in certain situations. Accordingly, it may not be enough that one simply hold a licensed self-conception; as I will discuss in the chapters that follow, actually attending to them as inputs in one’s deliberations may serve important ends.

Second, it is important to note that these principles are meant to be very general. I don’t take them to serve as exhaustive pronouncements of how one should think about one’s self-conceptions. To serve as a comprehensive guidance set, one

---

<sup>15</sup> Strictly speaking, value-incommensurate factive self-conceptions can be end-setting as well, in the sense that they give us reason to act so as to falsify them. For example, the true, value-incommensurate self-conception “Self-Centered Egoist” might recommend the end of not being a self-centered egoist any more. In stating this second principle as I have, I mean to limit the sense of “end-setting” to more positive (rather than “corrective”) ends, like helping to preserve the environment or helping your child learn how to read. I have accordingly covered the corrective ends that value-incommensurate self-conceptions provide in the third principle.

would have to supplement this list with further principles of greater specificity. Clearly, for example, Principle (4) cannot be the whole story of how one should manage ideal self-conceptions; given that there are potentially thousands of value-commensurate ideals one could pursue, it is probably not enough to simply say that one should acquire some of them.

This is not to say, however, that one couldn't give further guidance regarding roughly how many ideal self-conceptions one should acquire (e.g. more than five but less than 500) with what distribution (e.g. make sure you have at least one that covers each "deliberative sphere," such as family, career, relationship to country, etc.), or with what priority (e.g. *ceteris paribus*, one should acquire guiding ideals that are commensurate with a greater category of values than ones commensurate with a lesser category of values). That I am not endeavoring to provide more detail on these questions doesn't mean that one couldn't.

In terms of justification, I take these principles to be a logical consequence of the discussion up to this point. To provide something a bit more concrete, we might say that these principles are implied by our acceptance of MCDT, VBR, FBR, and perhaps the assumption that attending to self-conceptions in our deliberations is a good idea. If one accepts all of these claims, the practical principles enumerated above should come as no surprise.

What might Korsgaard or Frankfurt think of these principles? While it is unclear whether they would agree or disagree with Principle (1),<sup>16</sup> neither can fully account for Principle (2), since both are interested strictly in self-descriptions that the agent cares about. In terms of Principle (3), I don't think Frankfurt has the resources to endorse this claim, unless the value in question is wholeheartedness and the

---

<sup>16</sup> This is so because (as I discussed earlier) I'm not entirely sure whether either of them accepts FBR (although I suspect they might not).

disvalue in question is ambivalence. In Korsgaard's case, based on what she says about the Mafioso not having a reason to shed his practical identity of a Mafioso (despite the fact that it is in conflict with the value of humanity), we have good reason to think that she would not accept Principle (3) either, especially given the account of desire-independent value that I have been discussing. Similar points hold with respect to Principles (4) and (5); Frankfurt might be willing to accept them only insofar as the desire-independent value in question is wholeheartedness (which is a position that I have argued we should reject), and Korsgaard would have to reject them.

Since these practical principles have been stated in relatively "formal" terms, one might object that they aren't particularly useful. For example, telling me that I ought to hold only value-commensurate ideal self-conceptions (the objector might say) without giving me any advice on how to do this, or any specific guidance on which ideal self-conceptions are value-commensurate, isn't telling me much. In one sense there is something to this claim, but two comments are worth considering. First, these principles are arguably not quite as bereft of practical value as they may at first seem. As Aristotle has suggested, having a target at which to aim (perhaps even a somewhat "formal" target) may itself be a fairly important step toward the end of getting things right. Without having a clear sense of what one should be trying to do, the odds of going wrong are much higher. Perhaps if more people spent more time thinking about how their self-conceptions figure into their reasons for acting in the right ways, they would not only act on the right reasons with greater frequency, but perhaps their actions would make better sense to them.

Second, my aim in presenting these principles hasn't been to arm agents with tools that will guarantee successful deliberation, but rather to codify some of the conclusions that have come up in this chapter. These principles are in a sense simply

accompaniments to MCDT that have been provided to help flesh out some of its (admittedly general) practical implications.