CHAPTER ONE

INTRODUCTION

1.1 Project Overview

There seems to be something correct in the idea that there is a fundamental connection between our self-conceptions and our reasons for action. That I take myself to be a brother, a philosopher, and a rational animal of a certain sort seems to give me reasons for acting in certain ways. For example, when I deliberate about how to spend my weekend, that I hold the self-conception "Justin's Brother" seems to imply that I have reason to spend Saturday helping Justin find a job. Moreover, when I conclude that I have certain reasons for acting—reasons that I discover without thinking about my self-conceptions—these reasons nonetheless seem to say something about how I do or should think of myself. (Reflecting on the fact that I take myself to have reason to lie to others for my own gain, I may be led to wonder what this implies about me and the ways I think of myself. Am I really a selfish liar? Must I start thinking of myself this way?)

Reflections of this sort lead naturally to a number of questions. What precisely is the relationship between self-conceptions and reasons for action? (Where "self-conceptions" are understood broadly as descriptions that an agent either believes to be true of herself or is working to make true of herself.) Can we offer any law-like generalizations that capture the terms of this relationship? Are our intuitions about this relationship artifacts of the way that certain questions have been posed (since if one looks hard enough for some connection between two terms, one is eventually bound to find one), or do they reflect something more fundamental?

While I hope to answer all of these questions by this inquiry's close, it might be useful to start out with a more top-down view of the project I have in mind. Stated

at the highest level of generality, my aim will be to *make sense* of the relationship between self-conceptions and practical reasons. To achieve this aim, I will structure the discussion that follows around three central, organizing goals:

(1) To offer a precise and instructive statement of the relationship between self-conceptions and reasons for action.

(2) To describe the nature of this relationship (i.e. account for why it exists in the first place).

(3) To examine this relationship's implications and some of its more useful applications.

The first goal calls for a statement of this relationship that allows us to infer certain facts about practical reasons from facts about self-conceptions, and vice versa. If it turns out that a relationship only holds between a certain subset of self-conceptions and reasons for action, this statement will need to reflect this fact. As I will argue in later chapters, achieving this goal will yield a number of benefits, both theoretical (it will help us make sense of this relationship in a way that past attempts have failed) and practical (attending to this relationship in one's deliberations can prove quite valuable).

The second goal calls for an explanation or "accounting" of this relationship. If it turns out that this relationship is an artifact of the way that certain questions have been posed, we should be able to explain why this is so and why we are interested in the questions lead us to see this connection. If, on the other hand, this relationship is non-accidental—if there is an important sense in which self-conceptions and practical reasons are more fundamentally related—we should be able to explain why this is so.

The third goal calls for a discussion of the implications of this relationship. Once we have determined that this relationship is best stated in certain terms, the question arises: what does this fact about this relationship tell us about other things we

might care about? Can we use this statement of the relationship in our deliberations to our advantage? Similar questions arise with respect to the second goal. If we determine that this relationship is one of conceptual necessity, for example, can this knowledge help us answer other questions of interest?

It is worth noting that past treatment of these issues has been relatively light. In *The Sources of Normativity*, Christine Korsgaard has come closest to providing a sustained discussion of the ways in which one's self-conceptions impact one's practical reasons. Korsgaard's considered view is that there is a unidirectional relationship between the two, whereby all reasons for action are grounded in "practical identities," or self-descriptions that make an agent's life meaningful for him. As I will argue, this account leaves much to be desired. One may question both Korsgaard's claim that this relationship is limited to endorsed self-conceptions, and her claim that this relationship runs in only one direction.

After Korsgaard, Harry Frankfurt is the only other prominent philosopher who discusses these specific issues, but his treatment of them is very limited and indirect. What we get from Frankfurt isn't an explicit statement of the relationship, but rather a sparsely discussed and fairly inchoate view of it that is little more than a byproduct of other views that Frankfurt holds. This is not to say, however, that Frankfurt's account isn't worth examining. According to my reconstruction of his view, Frankfurt is committed to the view that this relationship is one of mutual determination, whereby any conclusions or discoveries that the agent makes with respect to one of these items (practical reasons or self-conceptions) will have a significant impact on the other. As in Korsgaard's case, I will raise a number of objections to this account, the most important of which concern its "internalistic" bent (like Korsgaard, Frankfurt limits this relationship to cared-for self-conceptions), its under-specification, and the fact that it treats this relationship in largely psychological rather than normative terms.

This leaves us with a field of a study populated by only two substantive views, both of which are problematic. Furthermore, not only do Korsgaard and Frankfurt fail to offer a fully satisfying account of this relationship, neither successfully explains why it exists in the first place (i.e. they don't even take up the second goal discussed above). Accordingly, while there are important lessons to be learned by looking at these accounts—both positive and negative—a fair swath of the discussion that follows will have to begin at ground level.

It is important to note that this charge of under-treatment holds only for the rather specific issues that I will be examining. There has been much discussion in the past on the relationship between what could be called the "shape" of practical reasons and the "shape" of the agents to whom they apply. For example, Thomas Nagel argues in *The Possibility of Altruism* that since agents have a certain shape (whatever else an agent happens to be, he is temporally extended and merely one agent among many), practical reasons must have a corresponding shape as well (so a given agent's reasons must be temporally extended and "objective"). Derek Parfit has discussed similar issues in *Reasons and Persons*, and more recently, Korsgaard has offered her own Kantian view on how the structure of agency demands that one's reasons reflect this structure (and vice versa) in her unpublished lectures on self-constitution.[1] These views on the relationship between agents and their reasons, however, differ from the more targeted questions that I have been raising, which concern the relationship between the *content* of particular self-conceptions and the content (and existence) of related reasons for action.

To give one illustration of why the project I am proposing is worth the bother—and to provide some background on what led me to take it up in the first place—my initial interest in these issues developed in response to an argumentative

---

[1] *Self-Constitution: Action, Identity and Integrity.*

strategy that I observed in number of moral theories. I noticed that many moral views—past and present—appeared to rely quite heavily on appeals to particular conceptions of what moral agents are like to justify their demands. Many moral philosophers, in other words, have attempted to justify their theories by convincing readers they "spoke properly" to true descriptions of what moral agents are like. While the adoption of this approach is more pronounced in ancient moral theories, it appears to be present in many modern theories as well.[2]

To give some examples, it is not uncommon to see claims of the following sort. Given that you are a rational animal of a certain sort, you ought to pursue activities that will develop and express certain moral and intellectual virtues. Given that you are a certain sort of rational agent, you ought to abide by the Categorical Imperative. Given that you are a temporally extended agent that can occupy both subjective and objective points of view, you should act only on the basis of "objective" reasons (since only objective reasons have purchase from both points of view).[3] While some of the moral views I have in mind don't rely on overt claims of this sort, they all seem to be at least tacitly committed to such claims. At the very least, there appear to be few moral views that one could not couch as appealing to a particular conception of what moral agents are like.

Furthermore, many of these theorists seem to assume that once the reader accepts a particular self-conception as a true description of what she is, there should no longer be any question of being motivated to act in accordance with the particular moral theory. We might sum up the central moves underlying these moral theories as an appeal of the following sort: you (the reader) have compelling reason to abide by

---

[2] Notable exceptions include moral theories that derive from claims about universal human desires (of the sort offered by Hobbes and Hume), and many forms of utilitarianism. While some conceptions of what agents are like are built into these theories, these conceptions don't play the particular sort of justificatory and motivating role that I have in mind.

[3] I offer these three claims as very rough glosses on the moral theories of Aristotle, Kant, and Thomas Nagel, respectively.

proposed moral theory M because you are a certain sort of thing (and once you see that you are such a thing—and hence come to hold the relevant self-conception—it will become clear why you have reason to abide by M).

It seemed to me that there was something right in this approach, but given the central role that this sort of argument plays in justifying many moral claims, it also seemed prudent to seek a better understanding of the presumed relationship between self-conceptions and practical reasons upon which it depends. Without a clear understanding of the terms of this relationship—much less some reason to believe that there is a valid relationship here to begin with—one might have reason to be skeptical of this entire justificatory enterprise. Moreover, even if one believed that there was a sound relationship here, one might wonder whether different accounts of it might rule out or lend support to particular moral theories.

This, in any event, should suffice as an overview of the project I am proposing and why it is worth pursuing. In Section 1.2 I discuss terminology that I will employ in the chapters that follow, and then conclude with an overview of these chapters in Section 1.3.

1.2 Terminology

As there is no consensus on how to define some of the terms that are central to this project, and since some of them are fairly technical, this section will be dedicated to terminology. We begin with the term "self-conception," which will serve as the genus for two particular species of self-description. My interest in distinguishing between these sub-categories of self-conceptions is based on the usefulness of doing so; I have chosen a taxonomy of self-descriptions, in other words, that maps conveniently on to a taxonomy (discussed at a later point) of the different roles that

self-conceptions can play in one's deliberations. Since there is no consensus on how these terms should be treated, the definitions that I will offer are stipulative.

Starting at the top, let us define a "self-conception" as any self-description that an agent believes to be true of himself or is working to make true of himself. This understanding of self-conceptions is meant to be quite broad; it includes not only "common" self-conceptions like "Brother," "Philosopher," and "U.S. Citizen," but also more specific self-descriptions like "Jazz Enthusiast," "5'10 White Male," and "Winner of the 1998 Gretzinger Award." The sum of one's self-conceptions should, on this reading, constitute what one might call an agent's comprehensive self-image, or reflective view of what she is and is working to become. As will become clear shortly, I am casting a fairly wide net with this term because I want to capture all self-descriptions of practical relevance, and as it turns out, "stock" and endorsed self-conceptions are not the only ones that fit this billing (even uncommon self-conceptions and ones the agent regrets can be reason-giving in certain contexts).

Based on this general definition, we can go on to sort all self-conceptions into two groups: *factive self-conceptions*—or descriptions that the agent believes to be true of himself (discussed in Section 1.2.1)—and *ideal self-conceptions*—or descriptions that the agent is working to make true of himself (Section 1.2.2). The importance of distinguishing between these two groups varies based on context; in some instances, the difference between them will have a large impact on how certain views are developed. However, in most cases I will be discussing self-conceptions more generally, simply because most of the claims that I will make hold true for both factive and ideal self-conceptions. In addition, I will also be discussing a separate category of descriptions that I will refer to as *attributed identities* (Section 1.2.3). Since attributed identities will be defined as descriptions that are attributed to an agent by someone else, but which the agent himself doesn't hold, I will discuss this category only to

distinguish it from self-conceptions proper. The final category of terminology that I will discuss concerns practical reasons and deliberation (Section 1.2.4).

1.2.1   Factive Self-Conceptions

Let us define a factive self-conception as any self-description that the agent believes to be true of herself. Given the wide scope of this definition, this class of descriptions will be rather broad; they include statements about one's relationship to others ("Justin's Brother," "American Citizen"), occupational titles ("Used Book Dealer"), cares and concerns ("Chocolate Lover"), physical make-up ("5'10 White Male"), etc.[4] While any given self-conception can be either true or false, all that matters when it comes to categorizing a description as a factive self-conception is that the agent *believes* it to be true.

It is important that we limit our use of this term to cases in which agents consciously (or at least semi-consciously) think of themselves under certain descriptions. In particular, it is important that we resist the temptation to collapse the gap between self-descriptions and the practical principles that are often associated with them (e.g. by claiming that someone holds the self-conception "Environmental Activist" simply because she cares about environmental issues).[5] While I don't want to deny that there is an important relationship between practical commitments and self-conceptions (a relationship that I will have quite a bit to say about shortly), collapsing one into the other raises serious problems, especially if one of our aims is to develop a clear statement of the relationship between self-conceptions and practical reasons. For reasons that will become clear in later discussion, there needs to be

---

[4] In contrast to Karen Jones' treatment, I don't seek to limit self-conceptions to those falling under "...the categories according to which we are accustomed to dividing people into kinds" (Jones: 76). I see no problem, for example, allowing the self-conception "Opera Buff."

[5] Korsgaard, Frankfurt, and Charles Taylor seem to flirt with this reduction at times.

enough of a gap between self-conceptions and practical commitments to ensure that the two can play different roles in an agent's deliberations.

One may worry at this point that this definition of factive self-conceptions is quite thin. All we know about them is that they are descriptions that agents think are true of themselves. Surely, however, deciding what is to count as a self-conception in particular cases is not always as straight-forward as this account implies. While the amount of detail we can add to this account will be limited given the nature of the subject (as Aristotle counseled, we should not seek more exactness than a subject allows for), it would be useful to consider three particular questions about what holding a factive self-conception requires. (Note that while I will raise and attempt to answer these questions with respect to factive self-conceptions, in many cases the same answers will apply in the case of ideal self-conceptions.)

First, one may wonder how convinced an agent must be that a given self-description is true for it to count as a factive self-conception. Demanding absolute conviction seems too strong; we might want to allow that an agent can hold a given self-conception without being convinced beyond any reasonable doubt that it is true. Perhaps, for example, descriptions that I take myself to have fairly good reason to fear or hope might be true of me should count as factive self-conceptions, whereas descriptions that I simply wish were true of me should not. Some degree of ambivalence on the agent's part concerning the truth of a given description should be permitted (how much, however, is a question that we might not be able to answer).

Second, one might wonder how consciously a given description must be held for it to count as a factive self-conception. This turns out to be a more difficult question to answer. Let us begin with two extreme cases, since they are fairly straight-forward. We clearly want to say that a given agent holds the factive self-conception S if she explicitly and sincerely avows holding it. At the other end of the spectrum, we

clearly want to exclude attributing a factive self-conception to agent if she explicitly and sincerely denies holding it. While these claims follow simply from our definition of a factive self-conception, there is also good theoretical reason to insist upon these requirements: If one of our central interests involves making sense of the role that self-conceptions can and should play in our deliberations, we must limit our discussion to self-descriptions that agents actually hold. A description that an agent doesn't consciously hold to at least some degree is a description that cannot figure as a premise in her deliberations.

Having ruled on these two extremes, however, there is also a vast gray area between them. What should we say about someone who, for example, spends several hours a week planning and implementing community betterment projects, but who has never consciously thought of himself as a "Community Activist" (or something of this sort)? As described, it would seem that we should not attribute this factive self-conception to this individual. But what if when asked whether he is a "Community Activist," this individual is quick to say that he is? After responding to this question it would be appropriate to attribute this factive self-conception to him, but what about before the question was asked? Should we take his quick avowal of this self-conception to indicate that he held a number of beliefs that, while perhaps not adding up to an explicit factive self-conception, were nonetheless close enough to count as one?

We can probably agree that when an agent is asked if he holds the self-conception S, if he is reluctant to admit that he does, or if it takes some convincing to get him to agree to the claim, we should probably conclude that this agent did not hold the self-conception in question (at least not prior to being "convinced," if he finally gives in). Trying to come up with an objective way of measuring reluctance, however, may be too problematic to serve as a criterion for settling these sorts of cases.

Another approach, which is perhaps more promising even if it is not without its problems, is as follows. Perhaps we can say that an agent A held self-conception S at time T if A was able to use S as a premise in her practical or theoretical reasoning at T. This account admittedly leaves several questions open (we still need to know what "able" means in this context), but it is one that I hope is at least potentially workable. The idea behind it is meant to track the general intuition that (for example), one test of whether I believe that Halloween falls on a Sunday this year involves looking at whether this proposition can serve as a premise in my practical or theoretical reasoning. Again, while this proposed criterion might not help us make a final ruling on the status of potential self-conceptions in many cases, it should at least indicate the sort of considerations that would be salient to making this sort of ruling.

Third and finally, there is basket of questions concerning how articulate the agent needs to be with respect to a given description for it to count as a self-conception,[6] and whether "images" rather than descriptions can count as self-conceptions.[7] While these are interesting questions, it isn't critical that we answer them given our central aims. It doesn't really matter where the line between "sufficiently articulate" and "insufficiently articulate" is drawn as long as what counts as a sufficiently articulate self-conception can figure as a premise in the agent's practical reasoning.

1.2.1.1 Sub-Categorizing Factive Self-Conceptions

While the definition of factive self-conceptions is quite broad, it is worth noting that there are several distinctions we can draw that can help us sort factive self-conceptions into further sub-groups. Since some of these distinctions will prove

---

[6] Owen Flanagan discusses related questions in his critique of Charles Taylor's in "Identity and Strong and Weak Evaluation."
[7] Karen Jones, for example, thinks that they will (Jones: 70).

useful in later discussion, they are worth identifying even at this early stage. In what follows, I will discuss three such distinctions.

The first concerns how easy it is to determine whether a given factive self-conception is true or false. In some cases, this determination will be fairly straight-forward. For example, we should have no problem deciding that the factive self-conception "Graduate Student" is true of me, and the factive self-conception "King of France" is false. On the other hand, there are cases in which determining the truth or falsity of a self-conception can be much more complex. Some descriptions, such as "Friend of Linda" may be open to interpretation, since there are no universal, clearly defined standards for self-conceptions of this sort. Take a case in which we are trying to determine whether A's self-conception "B's Friend" is true or not. Do A's attitudes toward B suffice to make true the claim that A is B's friend, or is the presence of reciprocal feelings on the part of B required to make it true? How strong do the feelings of each person toward the other have to be for A to truly be "B's Friend?" Are feelings alone sufficient? Do they need to be supported by the right sort of intentions and actions as well?

When the applicability conditions of a given description are vague in this way, it will often be appropriate to simply defer to the agent's own understanding of the self-conception. If A takes the self-conception "Friend" to be applicable to anyone who has certain feelings toward another person, we should allow that, if these conditions are met, it is true that A is "B's Friend." Note that this approach will only work in cases in which the agent herself has a relatively clear understanding of what a given self-conception requires. In many cases agents won't have given much thought to what it really means (or what they think it really means) to be an "American Patriot" or a "Good Samaritan." The point is simply that the applicability conditions for a given self-conception can range from "clear" to "quite vague," and where a given

self-conception lies on this scale might make an appreciable difference to how it is treated in certain contexts.

The second distinction that we can use to sort factive self-conceptions concerns the "valence" or "motivational import" that a given self-conception has for an agent. First there are self-conceptions that the agent finds motivationally compelling. Self-conceptions of this sort (which I will call "endorsed self-conceptions") include descriptions that the agent simply finds it pleasing to hold (e.g. "Witty Intellectual"), descriptions that the agent feels is of great practical importance (e.g. "Pacifist"), and descriptions that the agent feels obligated to act upon even if she isn't too keen on them. As an example of this last kind, consider someone who has been elected floor fire marshal by his co-workers on a day in which he hadn't come in to work. Despite resenting the fact that he is the "Floor Fire Marshall," this agent may nonetheless grudgingly concede the importance of the post, and hence be motivated to act as a "Floor Fire Marshall" ought to act. Roughly, if the agent is inclined to act in ways that are "recommended" by a factive self-conception, it will count as an endorsed self-conception.

Next we have self-conceptions that an agent regrets and is disinclined to express in her actions, such as "Social Outcast." Note that negatively valenced self-conceptions of this sort can still be reason-giving; as I will discuss in the next section, regretted self-conceptions may still be practically relevant under certain circumstances, although not in the same way as endorsed self-conceptions. Third, there are some self-conceptions toward which the agent may be utterly indifferent (e.g. "5'10 with Blue Eyes"). As in the case of regretted self-conceptions, I will argue that

this class of self-conceptions can practically relevant as well, despite the fact that the agent doesn't endorse them.[8]

The third and final distinction involves whether (and in what ways) a given self-conception may be essential to the agent's self-understanding. Here again we can identify three central categories. First there are factive self-conceptions that serve as essential descriptions of the *sort of thing* that the agent is (e.g. "Rational, Political Animal"). Second, there are self-conceptions that serve as essential descriptions of the sort of *person* who the agent is (e.g. "Pacifist"). Lastly, there are self-conceptions that are "accidental" from the agent's point of view (e.g. "Cornell Graduate Student"). The important distinction between these groups concerns what happens when a given self-conception is falsified. One the one hand, the passage of an event that falsifies a self-conception belonging to the first or second categories *could* be considered a threat to my very existence (although in two different senses). If the factive self-conception "Rational Animal" becomes false (perhaps I have been in a terrible accident), my very existence may be in jeopardy. On the other hand, I might be able to take in stride events that falsify "accidental" self-conceptions.

1.2.1.2 Extension Sets and the Reason-Giving Role of Self-Conceptions

Before moving on to ideal self-conceptions, we need to discuss one more term that will play an important role in the chapters that follow. This term finds its seat in the observation that many self-conceptions are closely associated with characteristic activities and concerns. The self-conception "Environmental Activist," for example, can be associated with the performance of environmental clean-up efforts, ensuring that one purchases only recyclable materials, etc. Many self-conceptions are in fact

---

[8] It is worth pointing out that I am parting company with most other writers on this subject (including Korsgaard and Frankfurt) in expanding the scope of this inquiry to include negatively valenced or indifferent factive self-conceptions.

characterized, and even constituted in large part, by activities and concerns of this sort.
To understand what an "Environmental Activist" is requires understanding the
practical principles that people falling under this description acknowledge.

Let us refer to the set of practical principles associated with a given self-
conception as that self-conception's "extension set." To illustrate, we might say that
the self-conception "Environmental Activist" is associated (either generally or only for
a particular person) with an extension set containing the following two practical
principles: (1) Contribute your time and resources to environmental protection and
clean-up projects; and (2) Use only environment-friendly products and dispose of
them in environment-friendly ways.

With this rough understanding of an extension set in hand, two qualifications
are worth mention. First, while nearly all of the self-conceptions that I will be
discussing have extension sets, not all of them do. Some self-descriptions aren't
characteristically associated with certain forms of behavior. The factive self-
conception "5'10 with Blue Eyes," for example, doesn't appear to have an extension
set. This is not to say that this description couldn't have an impact on one's
deliberations about how to act; for as it turns out, self-conceptions can figure into
one's practical reasons in one of two different ways.

On the one hand, since many self-conceptions are closely associated with
characteristic activities and concerns, they can *recommend practical projects*, as the
earlier "Environmental Activist" example illustrates. Accordingly, this self-
conception may figure into an agent's reasons by recommending that she help run the
upcoming Save The Lake rally. On the other hand, self-conceptions can also function
in one's deliberations as *practically relevant considerations*. For example, a local
restaurant in Ithaca offers half-priced meals to local waitstaff on Mondays.
Accordingly, the self-conception "Waiter" can prove practically relevant in certain

situations by figuring into an agent's thinking about where to eat. It is not part of our understanding of this self-conception that people with it ought to eat at Maxie's on Mondays (so it doesn't recommend practical projects in this case in the sense discussed earlier), but it does prove practically relevant.

The upshot is that while all self-conceptions can figure into one's reasons in at least one of these two ways (they can all be practically relevant in some contexts), not all of them recommend general practical projects as part and parcel of our very understanding of them. To return to my earlier example, the self-conception "5'10 with Blue Eyes" isn't typically associated with characteristic concerns or forms of behavior appropriate to this self-conception (and so it doesn't have what I am calling an extension set), although it can prove practically relevant in certain contexts (e.g. it might figure into my decision about whether to tryout for a part in a play calling for a tall, black actor).

The second qualification is that extension sets needn't be clear and articulate in all cases, either in terms of which practical principles they contain, or in terms of how these practical principles are formulated. In many cases the agent will have at best an inchoate understanding of what a particular self-conceptions requires of him. Nor is it the case that a given self-conception must have a fixed extension set associated with it. One's practical experience "living out" a given self-conception may lead to refinements of its extension set over time (as I'll discuss in Chapter 7). I have introduced the notion of an extension set merely as an idealized formalization of what I take to be an important component of our shared understanding of self-conceptions and the roles they play in our lives. (Note once more that while I have discussed the notion of an extension set in terms of factive self-conceptions, ideal self-conceptions have extension sets as well.)

1.2.2 Ideal Self-Conceptions

In contrast to factive self-conceptions, let us define an "ideal self-conception" as a self-description that the agent doesn't believe to be true of herself, but which she is working to make true of herself. For example, I might think there is value in being a "Considerate Friend," despite the fact that this description is not yet true of me. Perhaps I see that being inconsiderate is a bad thing, and yet I cannot help but lash out at my friends when I hear them complain about what I take to be very minor problems. In this case I hold the factive self-conception "Inconsiderate Friend" and the ideal self-conception "Considerate Friend." If my judgment is sincere, I should be working to make this ideal self-conception a true description of me (i.e. a factive self-conception).[9]

As with factive self-conceptions, we must limit the scope of ideal self-conceptions to descriptions that the agent consciously (or at least semi-consciously) applies to himself. In this respect, my notion of an ideal self-conception differs from what Amelie Rorty and David Wong refer to as an "ideal identity." On their view, "...a person...need not always be aware of her operative ideal identifications, and she need not always approve of those that are actually functioning" (Rorty and Wong: 24). I don't want to deny that one can be psychologically drawn to develop character traits without being aware of it (and perhaps against one's better judgment). Nor do I want

---

[9] While these two categories of self-conceptions are mutually exclusive, in some actual cases it could be difficult to determine into which category a given self-description falls. For example, if I am somewhat considerate, but not as considerate as I could be, is the self-conception "Considerate Friend" functioning in a factive or an ideal capacity? If this description is endorsed, it will be end-setting for me in either case. To keep things tidy, the best thing to do would be to insist that (a) every candidate self-description can serve either in a factive or an ideal capacity for a given agent at a given time, but not both, and (b) that these descriptions must have clear standards for their application (at least in theory). To return to our example, whether this description is factive or ideal will depend on whether its applicability standards are set above or below my present level of "considerateness." We could, for example, say that I hold the factive self-conception "Moderately Considerate Friend" and the ideal self-conception "Properly Considerate Friend."

to deny that one may, for example, take an interest in community-betterment projects without worrying about what sort of person one should be. It is important, however, for reasons discussed in the next section, that we distinguish between cases in which ideal self-conceptions play a direct role in the agent's reasoning *qua* ideal self-conceptions, and cases in which they don't.

To reiterate a point raised earlier, it should also be noted that many of the claims and distinctions I made with respect to factive self-conceptions apply in the case of ideal self-conceptions as well. For example, ideal self-conceptions are also characterized to a large extent by their extension sets. Similarly, the same criteria for determining whether an agent holds a given factive self-conception apply in the case of ideal self-conceptions. There are only two general differences between these categories of self-conceptions: (1) factive self-conceptions are thought to be true whereas ideal self-conceptions are thought to be false, and (2) while only some factive self-conceptions are endorsed, all ideal self-conceptions are endorsed.[10]

1.2.3 Attributed Identities

This brings us to our third and final self-description related term of interest. Let us define an "attributed identity" as a description that an observer attributes to an agent for the purpose of making sense of her behavior, but which the agent doesn't apply to herself. Attributed identities are basically a means for summarizing facts about an agent—including her desires, dispositions, capabilities, etc.—in a way that the observer thinks will help rationalize the agent's behavior. An attributed identity could be invoked either to explain a past action ("he's an environmental activist...of

---

[10] For an interesting discussion of one way in which ideal self-conceptions can figure into our self-improvement projects, see David Velleman's "Motivation By Ideals."

course he decided to donate the money") or recommend a future action ("as a community leader, she really ought to be more careful about the image she portrays").

Given the aims of this project, attributed identities are worth discussing for one reason only: namely, to distinguish them from self-conceptions so we can set them aside for the discussion that follows. Much of my interest lies in thinking about how self-descriptions can and should figure into an agent's assessments of his reasons for action. Attributed identities, however, are by definition incapable of playing this role since they aren't descriptions that the agent applies to himself.

Perhaps an example would help. Suppose that we say of someone "He is a community activist" in an effort to make intelligible his past work in support of the local community center. Furthermore, suppose that this agent has never thought of himself under this description. Perhaps he helped out in the community center in the past because he thought it would be a nice thing to do for his neighbors. Given this story, what should we make of the claim that "This agent helped out at the community center because he is a community activist?" Two possible interpretations present themselves, both of which are problematic.

First, if this claim is taken to be attributing a self-conception to this agent—i.e. if it is equivalent to: "This agent helped out at the community center because he thinks of himself as a 'Community Activist'"—then the claim is false. This agent doesn't hold this self-conception, so this is not an accurate explanation of what transpired. If, on the other hand, this claim is equivalent to "This agent cares about (or recognizes the value of) community betterment projects and that's why he helped out at the community center," then the claim will be true. However, on this second reading (which is an actual case of an attributed identity), self-descriptions aren't playing any role in the explanation. As the preceding gloss on this claim clearly shows, we can easily dispense with any talk of self-descriptions without losing anything; all the

speaker is really referring to is either the agent's interests, or the values that he recognizes in this action. Given the fact that this claim renders talk of self-descriptions unnecessary, we must exclude attributed identities from our discussion.[11]

1.2.4   Reasons and Deliberation

Having discussed the self-conceptions side of the relationship at issue, we must now look at reasons-related terminology. As I will be developing a detailed account of practical reasons in Chapter 4, I'll keep my comments here brief.

Following the run of literature on this topic, I will be using the term "reasons for action" (or "practical reasons") to refer to considerations that count in favor of performing a certain action. Within the general category of practical reasons, we can also draw a distinction between what have come to be known as motivating reasons and normative reasons. A motivating reason (as I'll be using the term) is a consideration that an agent thought counted in favor of certain action. While there is some disagreement in the literature as to what sorts of things constitute motivating reasons (the more traditional view identifies them with mental states, such as belief-desire pairs), [12] and without denying that beliefs and desires play an ineliminable role in any complete explanation of an agent's action, I will be identifying motivating reasons with the "foreground" considerations that agents consider in their deliberations rather than with "background" psychological states. [13] Regardless of where one lands on this question, however, everyone agrees that motivating reasons

---

[11] For an interesting case study of attributed identities (and the ways in which their persistent application can lead agents to accept them as self-conceptions), see Jean-Paul Sartre's *Anti-Semite and Jew*.

[12] Michael Smith, for example, favors the mental states view, whereas Jonathan Dancy favors a view similar to mine. See (Smith: 96) and (Dancy (1): 2).

[13] This decision is motivated by the terms of my inquiry. Since my focus throughout is on "foreground" considerations and deliberation (recall my claims about why we must set aside attributed identities, for example), I simply won't be spending much time talking about "background" psychological states (so there's little use in defining motivating reasons in terms of them).

are primarily explanatory. Motivating reasons tell us, in other words, why an agent performed a given action; they do not tell us whether the action was actually justified.[14]

Normative reasons, on the other hand, are considerations that actually count in favor of performing certain actions, and hence justify these actions. For example, while an agent might have thought that his hatred of his co-worker was a good reason to sabotage her upcoming promotion, it wouldn't be unreasonable to argue that this consideration wasn't a normative reason for this action (mere hatred of a co-worker cannot justify this act). Accounts vary on what determines the content of one's normative reasons (the usual suspects are actual desires, suitably informed desires, and values); I will remain agnostic on this question until Chapter 4.

So much for practical reasons. The final item of terminology to be discussed is "deliberation." It has become commonplace in the literature to use this term to refer to both (1) explicit chains of reasoning, and (2) thought processes that don't involve any explicit inferences, but which are taken to "model" or serve as approximations of concrete chains of reasoning. While this view leaves a number of questions unanswered, I will be following this standard treatment of what it means to "deliberate" in the chapters that follow.

Perhaps it would help to illustrate this point. Suppose I hear the smoke detector go off and so I rush to the kitchen. It will be unlikely that I have actually thought in great detail about what this sound indicates or demands. Let us assume, therefore, that I perform this action without thinking through anything like the following:

1. That beeping noise is the sound of the smoke detector going off.

---

[14] Note that motivating reasons can sometimes be presented as a justification of sorts when the agent is trying to excuse a past action ("I gave the man the bottle thinking that it was medicine and not poison"). Normative reasons, on the other hand, don't justify actions by excusing them.

2. The smoke detector usually only goes off when smoke is present.

3. The presence of smoke indicates that something is on fire.

4. There appears to be a fire in my house.

5. As a general rule, fires that spring up in one's house ought to be put out.

6. I ought to put the fire out.

7. To put the fire ought, I must go to the kitchen.

8. I must go to the kitchen.

Despite the fact that I haven't gone through any explicit deliberation of this sort, it is generally understood that, whatever actually went though my head prior to rushing off to the kitchen, it is acceptable to think of the above set of inferences as a "suitable expansion" of my actual thought process. In other words, we may assume that my actual thoughts modeled or approximated the inferences presented above, even if I didn't explicitly consider any of these premises.

In accepting this view, I am somewhat sympathetic to the account of deliberation that Paul Grice offers in *Aspects of Reason*. To help us determine what should count as "deliberation" or "reasoning," Grice proposes the following criterion: "...x reasons (informally) from A to B just in case x thinks that A and *intends* that, in thinking B, he should be thinking something which would be the conclusion of a formally valid argument the premises of which are a supplementation of A" (Grice: 16). The idea here is that we should count as "deliberation" any thought process—however truncated and informal it may be—that the agent intends as an approximation of a formally valid argument that could have been worked out in detail, even if this intention isn't explicit. Grice goes on to claim that it should be no great mystery why we often think about what to do in very inchoate and informal terms, and once we understand why this is so, it will become clear that his analysis of "reasoning" is correct.

> We have available to us (let us suppose) what I might call a 'hard way' of making inferential moves; we in fact employ this laborious, step-by-step procedure at least when we are in difficulties, when the course is not clear....Inferential judgments, however, are normally desirable undertakings for us only because of their actual or hoped for destinations, and are therefore not desirable for their own sake....Following the hard way consumes time and energy...and it would be, therefore, desirable if occasions for employing the hard way were minimized. A substitute for the hard way, the *quick way*, which is made possible by habituation and intention, is available to us, and the capacity for it (which is sometimes called intelligence, and is known to be variable in degree) is a desirable quality (Grice: 17).

In the pages that follow, I will count as "reasoning" any thought process on the agent's part that satisfies Grice's criterion. While a commitment to this approach leaves several questions unanswered, it should suffice for present purposes.

1.3 Chapter Overview

The progression of topics that I will pursue is based around the three organizing goals mentioned in Section 1.1 above:

(1) To offer a precise and instructive statement of the relationship between self-conceptions and reasons for action.

(2) To describe the nature of this relationship (i.e. account for why it exists in the first place).

(3) To examine this relationship's implications and some of its more useful applications.

My pursuit of the first goal will itself consist of three steps. I will first critique Christine Korsgaard's and Harry Frankfurt's accounts of the relationship in question in **Chapter 2** and **Chapter 3** respectively. After raising some initial objections to their views, I will proceed to the second step of this argument in **Chapter 4**, which contains my own account of what normative reasons for action are like. This chapter-length digression is necessary because (a) an adequate understanding of this relationship

requires an understanding of its terms (and it is normative—or "good"—reasons that will serve as our focus), and (b) there is a big question in the literature as to what normative reasons (and their existence conditions) are like. Against Korsgaard and Frankfurt, I conclude that normative reasons are grounded in desire-independent values rather than the desires of particular agents.

The third and final step towards this first goal comes in **Chapter 5**, in which I critically assess Korsgaard's and Frankfurt's views and build my own account of the self-conceptions/practical reasons relationship. After arguing that Frankfurt's view is superior to Korsgaard's in several respects—and hence that it would serve as a better starting point for a new account—I proceed to build upon this view in ways that both correct for noted problems, and incorporate the conclusions of my Chapter 4 account of normative reasons. The fruit of this labor is the Modified Co-Determination Thesis (MCDT), which maintains that the relationship at issue is one of mutual implication.

According to MCDT, facts about one member of this relationship (i.e. one's commitment to certain self-conceptions on the one hand, or the existence of normative reasons on the other) will in all cases have important implications for the other. For example, that you hold a certain self-conception (or more specifically, a true or "value-commensurate" self-conception) entails that you have associated normative reasons for acting. Conversely, that you have certain normative reasons for acting makes it appropriate for you to hold associated (true or value-commensurate) self-conceptions. One upshot of this view 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 self-conceptions, or his acting for this reason will have important implications for the appropriateness of holding certain self-conceptions.

My second organizing goal serves as the focus of the first half of **Chapter 6**. After reflecting on the nature of the self-conceptions/normative reasons relationship, I conclude that the connection between its members is one of conceptual necessity; we posses the concept of a normative reason in the first place because we need to figure out what to do, and conclusions about what to do must link us to actions that it makes sense for us (acting under particular descriptions) to perform. This view of the relationship is captured by the Agent-Addressed Reason Thesis (AAR), which maintains that every reason for action must be understood as addressed to a particular class of agents falling under a particular description. According to AAR, we should think of reasons for action as three-place predicates recommending (1) that a certain action be performed, (2) for the sake of a particular end, (3) by a particular sort of respondent.

In the second half of Chapter Six I begin working towards my third organizing goal by looking at some useful applications of AAR. In particular, I argue that properly attending to AAR's lessons can give us a new way to evaluate proposed actions, simplify our deliberations, justify the practice of grounding moral claims in appeals to self-conceptions, and help us make better sense of a number of other philosophical arguments.

My discussion of this third goal concludes in **Chapter 7**, which takes as its starting point the claim implicit in MCDT that there are two general patterns of deliberation that one can employ. On the basis of this view, I explore the possibility that using one of these patterns of deliberation—i.e. deliberating with the aim of *expressing* one's self-conceptions in one's actions—might help one realize values that are less available through non-expressive forms of deliberation. I conclude that the values of self-determination (understood as the value of determining what sort of self

to be) and self-knowledge are more available to agents who deliberate from self-conceptions to reasons for action.