CHAPTER SEVEN

THE VALUE OF EXPRESSING SELF-CONCEPTIONS IN ACTION

7.1 Introduction

The second half of Chapter Six explored a number epistemic and heuristic benefits that come from attending to the implications of the Agent-Addressed Reasons Thesis (AAR). These benefits all turn on the fact that AAR allows one to draw inferences from one side of a conception-action field to the other. The focus of this chapter is once again on the usefulness of thinking about the self-conceptions/normative reasons relationship, but in this case the benefits concern acting for certain reasons rather than reflecting on the implications of this relationship. More specifically, my focus will be on a particular pattern of deliberation that figured prominently in our earlier discussion of the Modified Co-Determination Thesis (MCDT), and the possibility that deliberating from self-conceptions to reasons for action (i.e. with the aim of *expressing* one's self-conceptions in one's actions) might put one in touch with values that are less available if one deliberates in non-expressive ways.

To set the stage for this discussion, it would be helpful to revisit MCDT in its final form. Recall that MCDT posits the following relationship between self-conceptions and normative reasons:

The Modified Co-Determination Thesis (MCDT):

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

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

As discussed in Chapter 5, one interesting implication 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 was prompted by consideration of his licensed self-conceptions, or his acting for this reason will make it appropriate for him to hold certain licensed self-conceptions.

Implicit in this view is the idea that we can distinguish between two different patterns of deliberation. One can either deliberate from a self-conception to a reason for action by thinking about what someone with the self-conception ought to do, or one can deliberate without considering self-conceptions at all. The second pattern of deliberation is by far the more common. Typically when we think about what to do, our deliberative gaze is firmly on the world of practical considerations and values. Nonetheless, there are occasions in which our interest lies in acting out, or perhaps "living up to," a certain self-conception.

Perhaps an illustration would help. Suppose that an agent holds the self-conception "Environmental Activist," and that the extension set of this self-conception contains only two practical principles, namely: (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. Suppose this person passes by a group raising money for a lake clean-up effort. We can imagine this agent employing two different patterns of deliberation to figure out what he should do. On the one hand, he might ask himself what an "Environmental Activist" should do in these circumstances, and conclude that he should donate money *because* this is what an "Environmental Activist" should do in this situation. In this case, the agent has deliberated from a self-conception to a reason for action. For present purposes it doesn't matter whether this person (a) explicitly considered the practical principles in the relevant extension set (qua principles associated with this self-

conception) to figure out what an environmental activist should do, or (b) reflected more abstractly upon what environmental activists are like (perhaps picturing an environmental activist in action itself suggested an appropriate response). In either event the ultimate goal of this agent's deliberation—what controlled the process of deliberation that he went through—was figuring out what someone with this self-conception ought to do in these circumstances.

On the other hand, this person's deliberative gaze might have been focused entirely on relevant practical principles (although in this case assume that he didn't consider these principles qua their association with this self-conception). He might have simply thought: "it is a good idea to contribute to worthy environmental projects (one should always do so when one gets the chance), and this is just such a project." In this case, the agent is not at all concerned with self-conceptions, but merely with practical principles that he endorses. The ultimate goal of his deliberation in this case *is not* to figure out what someone with a particular self-conception should do.

With a rough understanding of these two patterns of deliberation in hand, note that I will be making a simplifying assumption about them in what follows. I will be assuming that similarly situated agents who deliberate about what to do by using these two deliberative patterns will always arrive at the same practical conclusions. In other words, if we imagine two agents who are alike in every respect except that agent A always deliberates for the express purpose of acting out his self-conceptions, whereas agent B always deliberates without regard to his self-conceptions, I want to assume that A and B will always draw the same conclusions about which outward actions to perform. (We need to limit this assumption to "outward actions" because there is an important sense in which these two agents can never perform precisely the same actions. If we take actions to be individuated by the descriptions under which they are performed, since these agents will always perform their actions under different

descriptions, they will never perform the "same actions.")[1] Moreover, I will assume

this despite the fact it is unlikely that A and B would always make the same choices

about what to do.

If this assumption is fairly unrealistic, why make it in the first place? Two

points. First, it seems unlikely that anyone could ever prove that one of these patterns

of deliberation more reliably leads to the "right answer" in a majority of cases. One

pattern might work better in some situations, while the other might work better in

others. Second, my main interest is whether one of these patterns of deliberation has

what we might call "beneficial side-effects;" i.e. whether deliberating in one way puts

the agent in touch with values that are independent of actually drawing the correct

conclusion about what to do. Accordingly, while one pattern of deliberation might

lead to the right conclusion in a particular sort of situation more reliably than the

other, since there are no general facts about which pattern is more reliable on the

whole, and given that my interest requires abstracting from reliability considerations

anyway, let us assume that the two patterns of deliberation will always recommend

performing the same outward action.[2]

As for why this inquiry is worth the effort, one reason should be quite obvious.

If it turns out that certain values are more open to agents who deliberate in one way

---

[1] One may wonder why we should bother speaking of the "same outward actions" in the first place if this is so. If actions performed under different descriptions are really different actions—and if the descriptions in question often make a big difference to our assessments of these actions—is there any sense in pointing out these sorts of outward similarities? In some contexts, the answer will be "no." There may be little point in claiming that two outward acts of jumping in a lake to save a drowning person are the same if one was performed under the description "save that person" and the other was performed under the description "bring him ashore so I can continue torturing him." In the sort of cases I have in mind, however, our assessments of the actions in question will probably overlap significantly. The environmental activist example is a good case in point. This is not to say that this worry is groundless even given my focus; I will return to it in the final section of this chapter.
[2] The assumption that neither pattern is more generally reliable than the other is an important one. If it turned out that one pattern was a better bet on the whole, then my interest in abstracting from whether a given pattern "gets things right" would be moot; who cares if deliberating one way puts one in touch with, for example, the value of self-determination, if in the process it reliably leads one to perform the wrong actions?

rather than another, this would be very useful information to have. Beyond this consideration, however, it is also worth noting that the answer to this inquiry bears on our earlier discussion of the practical principles cataloged in Chapter 5 (which included guidance on what sorts of self-conceptions one should hold and how one should treat them, etc.). For if it turns out that deliberating from self-conceptions to reasons for action had nothing to be said for it, there might be less interest in developing these practical principles in the first place. These principles will of course be valid whether or not the "interest" in question can be established, but the value of attending to them in one's deliberations would be doubtful. Without having some reason to deliberate from self-conceptions to reasons for action, one could (as one does most of the time) go through much of one's life without ever attending to self-conceptions without any real loss. Accordingly, at stake in this chapter is also a reason for attending to the practical principles discussed back in Chapter 5.

To sum up, this chapter's guiding aim will be to determine whether there are values that are more open to agents who deliberate from self-conceptions to reasons for action. In the interest of simplifying the terms of this inquiry—and to situate it alongside a set of related issues—I propose that we think of this question in terms of the value of "expressing" one's self-conceptions. We might wonder, for example, whether there is more value in donating money to the lake clean-up project as an expression of a self-conception than in donating the money on some other (albeit related) ground. Given the ambiguity of the term "expression" (initial questions about its application surfaced in Chapter 2 discussion of Korsgaard's view), I will begin this chapter by defining a particular sense in which an act can be understood as expressing an agent's self-conceptions (7.2).

After developing this account of expression, I will examine a number of arguments for the claim that it is valuable to deliberate in ways that express one's self-

conceptions. The first set of arguments holds that the value of self-determination is more open to agents who deliberate in ways that express their self-conceptions (7.3 and 7.4). After concluding that one of these arguments has some promise (by showing that one may be able to better determine what sort of "self" one is by engaging in expressive deliberation), I will explore an argument that links expressive deliberation to the value of self-knowledge (7.5). I conclude by exploring the possibility that there might be some disvalues associated with expressing one's self-conceptions in action (7.6).

7.2 "Agent-Expression"

In this section I will describe the conditions under which an action can be said to express a self-conception. The resulting definition will allow us to answer questions of the following sort: if I help my brother find a job, what has to be the case for this action to count as an expression of my self-conception "Brother?" The view that I will defend holds that there are two necessary conditions for an action A to express a self-conception S. (Note that in what follows, I will be limiting discussion to self-conceptions that are characteristically associated with certain sorts of behavior.)

First, there is what I will call the *extensional requirement*, which demands that A be recommended by members of S's extension set (i.e. the practical principles that are associated with S). The reason behind this requirement should be fairly clear. Think back to our environmental activist example. Assuming that we properly identified the members of the relevant extension set, it is hard to see how playing checkers for hours on end could be an expression of the self-conception "Environmental Activist" (at least not without telling an elaborate story linking a game of checkers to the promotion of sound environmental practices). Lacking a story of this sort, it seems false to claim that playing checkers is expressive of this self-

conception. The extensional requirement simply reflects the idea that most self-conceptions are associated with (and partially constituted by) characteristic forms of behavior.

Second, there is what I will call the *intentional requirement*. While this condition is more difficult to articulate than the first, it will prove quite important for the discussion that follows. The intentional requirement holds that, in addition to A's being recommended by a member of S's extension set, A must be recommended by a member of S's extension set *qua member* of S's extension set. The intentional requirement, in other words, requires that A be understood (by the deliberating agent) as "giving voice" to S.

To see why this is so, note that every intentional action must be performed under a given description. The description under which an agent acts is what makes an action the particular action that it is. To borrow an example from Elizabeth Anscombe, that I intend to move the pump handle up and down under the description "fill this bucket with water," makes it the case that my action is one of filling the bucket with water rather than one of tapping out a certain rhythm with the pump handle.[3] This doesn't mean that every time an agent performs an action she has an explicit and articulate description of what she is doing in mind, but there must be some description that adequately captures the agent's intentions if there is to be a genuine action at all.

Returning to the intentional requirement, the idea is that the description under which an agent performs an act has to make the right sort of reference to a self-conception if it is to count as an expression of this self-conception. More specifically, the agent has to understand the action as worth performing because it is recommended by a particular self-conception.

---

[3] See (Anscombe: 37-8).

To see why we need this second requirement, consider what happens if we leave it out of our account of expression. Suppose that John and I work for the same company and we are each working on different sections of a group report. I see that John will have to stay late to get his part done on time, and so I volunteer to stay with him to help him finish. Supposing that I have not decided to perform this action as an explicit expression of a given self-conception, a question arises if we reject the intentional requirement: precisely which self-conception have I "expressed" by performing this action? Suppose my thinking in this case was: I should stay late to help John finish his report so he can get home at a reasonable hour (this is the description under which I have performed this action). In performing this action, have I expressed the self-conception "John's Colleague," or "John's Friend," or "Good Samaritan," or something else? A good case can be made the claim that I haven't really expressed any of these self-conceptions. Perhaps all I have "expressed" is my willingness to help John.

To generalize, the problem here is that without the intentional requirement, one could argue that any given act—even one performed on the basis of a rather detailed maxim of what one proposes to do—may be potentially "expressive" of many different potential self-conceptions (namely any that satisfy the extensional requirement). To really drive this point home, we could couch this problem as one of "projectability." Just as any set of data points could be explained by an infinite number of scientific theories, any given set of actions that are not linked by one's intentions to a specific self-conception could be considered "expressive" of a potentially infinite number of possible self-conceptions. This problem can be obscured by our tendency to focus on "stock" self-conceptions, like "Brother" and "Community Activist," but there is no theoretical prohibition against holding more complex and uncommon self-conceptions. Surely we don't want to claim that a given

act expresses every self-conception that could possibly recommend it. The problem is that we need to come up with a way of determining which of the potentially infinite self-conceptions are actually expressed by a given action, and the only viable solution appears to be this: we can determine which self-conceptions an action expresses by looking at which self-conceptions figured into the agent's deliberations.

It is worth mentioning that a rejection of the intentional requirement also runs afoul of one of the guiding constraints of our broader project. Back in Chapter 1 I drew a distinction between self-conceptions and attributed identities, and argued that we would need to focus exclusively on the former because the latter—by definition—cannot figure into an agent's deliberations about what to do. The implication of this fact is that attributed identities cannot explain an agent's action in the way that someone who is interested in making sense of first-person deliberations would want.

To illustrate this, consider the following example. Suppose I try to explain an action on your part by saying: "he's a community activist...of course he doesn't mind spending his weekends cleaning-up the neighborhood." Let's assume that you don't hold the self-conception "Community Activist." There are two possible interpretations of the explanation I have just offered, both of which are problematic.

First, in claiming that you are a "Community Activist," I might be referring simply to certain motivational tendencies on your part (so all I really mean is: "he tends to go in for community clean-up projects, so this action is no surprise"). On this reading, however, possible self-conceptions aren't really figuring into my explanation at all, so there is no sense in speaking of self-conceptions in the first place. Second, I might be genuinely attributing to you a self-conception that we have just stipulated that you don't hold. In this case, my explanation seems to have misfired. I have basically claimed: "he performed this act because he takes himself to be a Community Activist," which is false given the way that this example has been described.

A similar problem faces us if we reject the intentional requirement. Suppose we insist that an action is expressive of a self-conception that the agent either doesn't hold or didn't consider in her deliberations. In making this claim, we will either be misspeaking in talking about self-conceptions in the first place (since all we are really referring to is a motivation tendency, which means we could eliminate any reference to self-conceptions), or our claim will be false (since the description in question is an attributed identity, which cannot figure into an adequate explanation of what happened).

Based on these considerations, it seems as though the most one can conclude in our earlier "stay late to help John" case is that I have "expressed" my acceptance of a practical principle calling on me to help people in John's position. This understanding of what transpired, it will be recalled, is similar to the one that G.A. Cohen discussed in his critique of Korsgaard. There he suggested that: ...[it is trivial that] if I treat something as a reason, then it follows that I regard myself as, identify myself as, the sort of person who is treating an item, here and now, as a reason. I do not see that [endorsing an action in this way] must issue in such a [self-]conception, other than in the indicated trivial sense" (Korsgaard (1): 185). Unless the action is linked to a particular self-conception in the agent's deliberations, we risk facing a situation in which any given action can be expressive of a large number of things, none of which figured into the agent's decision. The conclusion is that the sense of expression that we are after—which I will hereafter refer to as "agent-expression" to distinguish it from other senses of the term[4]—requires the addition of the intentional requirement.

---

[4] Since common use of the term "expression" is somewhat varied, I will accordingly use the term "agent-expression" both to rule out possible confusions, and to indicate that I don't take the foregoing discussion to constitute an analysis of the concept of "expression." Based on common usage, I don't want to deny that many people will feel quite comfortable speaking of an agent "expressing" self-conceptions in situations in which it is clear that the agent hasn't "agent-expressed" any self-conceptions.

To help make better sense of this notion (i.e. to go beyond the mere claim that agent-expressive deliberation has to meet the intentional and extensional requirements), let us say than act is agent-expressive if the agent deliberated *with the goal* of figuring out what a person with the self-conception in question should do in certain circumstances. More specifically, we could say that in cases of agent-expression, this particular goal will determine: (a) how the deliberation proceeds, (b) at what point the agent concludes that she can stop deliberating (because the purpose of the inquiry has been achieved), and (c) whether her deliberation has been a success. I express a self-conception in the requisite sense when I understand my action to "issue from," or be legitimized by, the self-conception. The action that I settle upon, in other words, will be the right one to perform *because* it is the sort of action that someone with the self-conception in question should perform.

This is not to say that determining when an act of agent-expression has occurred will be easy. One may wonder how explicitly one needs to consider self-conceptions for there to be an actual act of agent-expression. This is a difficult question, but this sort of problem is common to any view that ties any attribution conditions to the content of actual deliberations. Keeping in line with my treatment of this issue in Chapter 1, let us say somewhat nebulously that an act is agent-expressive if a "suitable reconstruction" of the description under which the agent acted meets the criteria presented above. As it turns out, given our present aims it isn't critical that we come up with a clear test for settling tough cases. All that really matters is that some acts will be clearly agent-expressive and some acts clearly won't. As long as both patterns of deliberation are available to agents (as they clearly are), it will be worth our while to ask whether there is something to be said for acting in agent-expressive ways. The remainder of this chapter will be dedicated to this very task.

Before turning to answers to this question, however, it is important to note that the discussion that follows will be limited to a certain subset of self-conceptions. We can distinguish between (1) self-conceptions the truth of which depends on an agent's acknowledging and expressing the practical principles associated with it, and (2) self-conceptions that count as true or false of an agent regardless of whether she acknowledges and expresses associated practical principles. A clear example of the former is "Environmental Activist." If an agent doesn't think she has reason to protect the environment or recycle and doesn't perform these activities, this self-conception will be false. A good example of the latter is "Diabetic." While this self-conception might be associated with characteristic activities (monitoring blood-sugar levels, etc.), it will serve as a true or false description of an agent regardless of whether she acknowledges or performs these activities. This distinction is worth noting because even though one can agent-express both sorts of self-conceptions, the "beneficial side-effects" I'll be discussing in what follows apply only to type-(1) self-conceptions.

7.3 First Hegelian Argument From Self-Determination

I will begin by examining two Hegelian arguments for the claim that one is more likely to realize the value of self-determination if one acts in agent-expressive ways. As I will discuss, each argument involves a different sense of what it means to be self-determining. The first concerns the value of determining for oneself what actions one will perform, and the second concerns the value of determining what sort of self one will be. At the discussion's close, I will conclude that the second argument holds more promise than the first.

Before looking at these two arguments, however, two preliminary points are in order. The first involves the sense in which these arguments are "Hegelian." In the discussion that follows, I will be relying quite heavily on Fredrick Neuhouser's

discussion of Hegel in Chapter Three of his *Foundations of Hegel's Social Theory*, for the simple reason that Neuhouser manages to present a readily accessible account of the (often obscure) arguments in question. Accordingly, while I will refer to the following arguments as "Hegelian," my starting point is really "Neuhouser's Hegel."

Furthermore, my own reconstruction of Neuhouser's arguments will be tailored in some instances to reflect present interests. My discussion of the Hegelian arguments that follow will accordingly occur at two removes from Hegel's own discussion; not only will I rely largely on Neuhouser's exegesis, but I will also be making changes to this exegetical base at times to suit my own ends (most notably in terms of splicing off various strands of what I think is supposed to be a unified argument). We might accordingly characterize the following arguments as "Hegelian inspired."

Second, the focus of this discussion will be on what Neuhouser refers to as "practical identities," which are roughly descriptions of the particular roles that one occupies within social institutions (Neuhouser: 94). This is important to note because many of the claims that follow are meant to apply only to a limited subset of self-conceptions (roughly: endorsed, value-commensurate self-conceptions that specify how one is related to particular social institutions, like the family, civil society, etc.). With these preliminaries in hand, let us begin with the first Hegelian argument, which can be summarized as follows:

First Hegelian Argument for Self-Determination:

1. <u>Freedom as the Highest Value</u>. The highest object of value for human beings is practical freedom, understood as self-determination (humans are essentially self-determining creatures).

2. <u>Appropriate Sense of Self-Determination</u>. On this view, I am self-determining to the extent that my actions issue from my own will.

3. <u>Self/External World Relationship</u>. Since my actions "take place" in the external social world, I face the persistent possibility that my will might be actively resisted by this external social world. As a result, I face the persistent threat of being constrained and determined by the external social world, which may impose upon me laws and principles of behavior that I cannot accept (i.e. which are opposed to my will).

4. <u>Role of Practical Identities</u>. Practical identities represent solutions to the self/external world relationship that threatens my ability to be self-determining. Practical identities act as a bridge through which I interact with the external social world and make it possible for my will to be continuous with it. By acting under the guise of a practical identity—like "American Citizen" or "Teacher"—I can see my will executed out in the social world amidst social institutions that are receptive to (and hence continuous with) my will. The result is that I no longer see the social world as a check on my will, but rather an extension of it. "By participating in the institutions of *Sittlichkeit*....social members find in the external, social world a confirmatory reflection of who they take themselves to be..." (Neuhouser: 107). A practical identity is essentially a self-description and a set of associated activities that brings me (under a given description) and the external world (understood as a hospitable forum for my willed actions) together, thus resolving the threat of external constraint.

5. <u>Conclusion</u>. It is through holding a practical identity and expressing it in my actions that freedom as self-determination is achieved.

For example, the practical identity "U.S. Citizen" serves to connect me to the larger social world that I inhabit. It is through this self-description that I see myself "plugging into" this larger social world, in part because this social world is constructed so as to receive this practical identity. When I act in ways expressive of

this practical identity, I don't experience my actions as coming up against something that is hostile to my will. To put this point in more Hegelian terms, I come to see this social world as a home for me, a place that is continuous with my willed actions.

This is a rather interesting argument in several respects, but it is not without its problems. First, one might worry that this argument as a whole is a bit too abstract. There seems to be something in its talk of reconciling conflicts between one's will and the theater in which it is expressed, but thinking of concrete examples that illustrate the problem that practical identities are supposed to solve turns out to be rather difficult. Furthermore, to the extent that this argument works at all, it only works for a relatively small class of possible self-conceptions: namely those that locate one with respect to central social and political institutions.

A more serious worry given our present aims, however, is that achieving freedom as self-determination doesn't actually require that one act in agent-expressive ways. It appears I can plug in to the larger social world and freely decide what to do without thinking about my practical identities at all. Why can I not be self-determining in the requisite sense simply by deliberating from practical principles to reasons for action, so long as the actions I endorse conform to the laws and social rules of the institutions to which I belong? To achieve the sense of self-determination at issue, all that is required is that my will be "continuous" with the social world, which doesn't itself require any explicit thoughts about my self-conceptions.

This point alone is sufficient to remove this argument from the purview of our inquiry. Note, however, that this objection isn't meant to undermine the more general Hegelian argument at issue. Since Hegel was not out to demonstrate the rationality of performing acts of agent-expression, this is not necessarily a problem for his view.

7.4 Second Hegelian Argument From Self-Determination

      This brings us to the second Hegelian argument for self-determination, which runs as follows:

1. <u>Freedom as the Highest Value</u>. The highest object of value for human beings is practical freedom, understood as self-determination (humans are essentially self-determining creatures).

2. <u>Sense of Self-Determination</u>. On this view, the sense of self-determination at issue concerns the more literal determination of what sort of self one is to be (I am self-determining to the extent that I determine who I am).

3. <u>Determining Who You Are Requires Expressing Your Practical Identities</u>. Simply holding a given practical identity isn't sufficient for self-determination; the freedom at issue is practical rather than theoretical, and so is the form of self-determination. What determines whether a given practical identity will count as a true description of me is not merely my endorsement of it, nor simply "passive" facts about relations that I have to the world (e.g. having the same parents as someone else), but rather my activities. I actually become a brother, citizen, etc. by carrying out the activities that these roles require.

4. <u>Conclusion</u>. It is largely through holding practical identities and expressing them through my actions that I become self-determining.[5]

      Here's an example. Suppose that I hold the self-conception "American Patriot." One might argue that simply holding this self-conception isn't enough to make it true; I would actually have to act in ways expressive of this self-conception to give it the proper legitimacy (e.g. I might actually have to lobby for some cause that I took to be "pro-American"). Accordingly, it may be the case that I cannot fully

---

[5] It should be noted that the two Hegelian arguments offered so far aren't necessarily independent; they do overlap in some respects, and draw on some of same intuitions.

determine myself to be an "American Patriot" unless I act in ways expressive of this self-conception. Neuhouser remarks:

> As members of *Sittlichkeit*, they do not merely imagine themselves to be mothers, plumbers, and citizens; they in fact become such beings by carrying out the activities that the roles of mother, plumber, and citizen require....they give objective reality to their otherwise merely subjective identities. Thus, the social participation of the members of *Sittlichkeit* can be said to have a self-expressive character, insofar as it is an activity through which they externalize, or give expression to, their conceptions of who they are (Neuhouser: 107-8).

Neuhouser adds, "The type of self-determination carried out here could be characterized as an act of *giving real determinacy to the self*—and hence as a kind of *self-realization...*" (Neuhouser: 109).

At this point we must again determine whether one must perform acts of agent-expression to achieve this form of self-determination. Why can't one give "real determinacy" to oneself as an American Patriot simply by acting in accordance with (what turns out to be) the extension set of "American Patriot," regardless of whether one does so *qua* recommendation of this self-conception's extension set? Why is an act of agent-expression needed for one to be fully self-determining in this sense?

To answer this question, let us return to the John's Friend/Colleague example discussed back in Section 7.2. Recall that since the act of staying late to help John is recommended by the extension sets of a number of different self-conceptions, we must conclude that if I don't consciously perform it as an expression of a particular self-conception, it is unclear which (if any) self-conception it expresses. The corollary point in the case of self-determination is this: if we insist that I have determined myself to be something by staying late at the office to help John, it will be equally unclear what precisely I have determined myself to be. At best, we might say that I have determined myself to be the sort of person who stays late at the office to assist

co-workers who are working on related projects. But there is no more *general* description of myself (of the sort that we typically offer when discussing self-conceptions, such as "John's Friend" or "Good Samaritan") that I have made true of myself. In terms of the potential self-conceptions at play in this example, since I haven't performed an act of agent-expression with respect to any of them, nor have I determined myself to be someone for whom any of them are true.[6]

At this point, however, one might wonder whether this really matters. Suppose that in this case I have "merely" determined myself to be the sort of person who stays late at the office to assist co-workers...etc. This is still an instance of self-determination, isn't it? If so, why claim that acts of agent-expression help one become more self-determining in the appropriate sense?

Broadly construed, I think we should allow that this is an act of self-determination. This is why we should speak of the ways in which agent-expression might make one *more* self-determining. The thesis I would like to consider is (more narrowly) that acts of agent-expression can affect the depth and scope of one's ability to be self-determining.

How so? Since most "stock" self-conceptions are associated with many characteristic forms of behavior (i.e. their extension sets can be quite large), coming to hold a licensed self-conception of this sort commits one to acting in a variety of ways. Suppose that I stay late to help John finish his report by deciding that this is what a true friend would do in this situation. This act of agent-expression is also an act of

---

[6] Note that this suggestion isn't at odds with the MCDT claim that in recognizing certain normative reasons for action, it will be appropriate for me to hold certain self-conceptions. The question of whether this act of staying late at the office makes it appropriate for me to think of myself as "John's Friend" or "John's Colleague" is quite different from the question of how I came to the conclusion that this reason exists in the first place. It is, in other words, open to me to decide that I ought to help John without considering antecedent self-conceptions, and conclude (perhaps several days later) that recognizing this reason implies that I ought to think of myself a certain way. In this case—depending upon how particular past actions are specified—we could say that this normative reason makes it appropriate for me to hold one of several self-conceptions, including "John's Colleague," "John's Friend," or "Good Samaritan."

self-determination in the sense that it either helps make it the case that I am "John's Friend," or it reinforces this pre-existing fact about me. The difference between determining myself to be "John's Friend" rather than determining myself to be "the sort of person who stays late at the office to assist co-workers who are working on related projects," is that the former implicates me in many more practical commitments than the latter. In this latter case, perhaps all I have committed myself to is helping other co-workers in very similar circumstances.

Determining myself to be "John's Friend," on the other hand, has a much greater impact on what sort of person I become. Not only should I stay late at the office to help John out, I should also help him move furniture, celebrate this birthday, confide in him, etc. To give another example, if I stay late to help John as an explicit expression of the self-conception "Good Samaritan" (thus perhaps either determining myself to be a Good Samaritan or reinforcing a previous commitment to this self-conception), I will open myself up to a number of rather different practical demands. Not only will I have determined myself to be the sort of person who helps co-workers in these sorts of ways, I will also have determined myself to be the sort of person who helps stranded motorists, assists old ladies across the street, etc. Proceeding through life without performing any acts of agent-expression might accordingly limit one to a very piece-meal process of self-determination, and can potentially leave much of one's practical existence relatively "undetermined."[7]

It is also worth pointing out that, since self-conceptions are more than just sets of practical principles, the sense of self-determination at issue when one agent-expresses a robust self-conception of this sort goes further than simply committing oneself to a number of practical principles. Many self-conceptions have a certain

---

[7] This point bears some affinities to the point I made in Chapter Six about the "deliberative economy" that results from focusing on self-conceptions when trying to figure out how to act. Both points turn on the fact that many self-conceptions encapsulate large collections of practical principles.

"feel" to them as well; if I determine myself to be an "American Patriot," for example, I may be also signing on to a particular image charged with a certain emotional resonance, a sense of fellowship with other citizens, a new understanding of myself as fitting into a tradition of American patriotism with a rich history, etc. While practical principles are perhaps the most significant components of many robust self-conceptions, they are certainly not the only components.

The upshot is that acting in agent-expressive ways can make one more self-determining (or may at least make one a more efficient self-determiner). At this point, however, we need to qualify this claim in two respects. The first qualification concerns a possible alternative route to satisfying the intentional requirement. So far I have been speaking of acting in agent-expressive ways in response to a stretch of deliberation in which the agent moves from a self-conception to reasons for action. Reflecting back on the implications implicit in MCDT, however, it appears as though the process of linking self-conceptions to reasons needn't come at the time of action.

For example, suppose that at time T1 I decide that I should help John finish his report without considering any self-conceptions (a "suitable reconstruction" of the description under which I perform this act makes no reference to my self-conceptions). The resulting action is accordingly not agent-expressive. Now suppose that at time T2, after thinking about MCDT and its implications, I realize that my recognition of this reason makes it appropriate for me to hold certain self-conceptions. (Perhaps this action, in conjunction with other past acts, makes it appropriate for me to think of myself as "John's Friend" or a "Good Samaritan"). It could be argued that I have at T2 met the intentional requirement (in a somewhat loose sense) by conceiving of this action as a basis for holding a particular self-conception. We may accordingly say that while the T1 act wasn't agent-expressive (and so I wasn't "robustly" self-determining at T1), my T2 reflections are agent-expressive, and hence my realization that it is now

appropriate for me to hold the self-conception "John's Friend" (along with my subsequent commitment to this self-conception) makes it the case that I am robustly self-determining at T2.

How does this bear on the view in question? Perhaps we should update it in the following way: an agent will be more self-determining to the extent that she (a) acts in agent-expressive ways (i.e. explicitly deliberates from self-conceptions to reasons for action), or (b) comes to think of past actions as a basis for holding certain self-conceptions by engaging in what we might call "agent-expressive reflection." The only difference between the two cases concerns when the act of self-determination occurs. In both cases one becomes more self-determining when the intentional requirement is met, but in (a)-type cases this happens concurrently with the initial deliberation, whereas in (b)-type cases it happens after the fact.

The second qualification comes in response to the objection that this account is too austere. Suppose that I spend quite a bit of time with Susan. We go to movies together, share secrets, look out for each other, etc. Furthermore, suppose that I have never explicitly thought of myself as "Susan's Friend." Unrealistic as it might be, let us further assume that I would be surprised to hear this label applied to me. According to the account just offered, it appears as though I haven't actually determined myself to be "Susan's Friend." One may worry, however, that this is a counterintuitive result.

In response, I think we should agree that I have indeed determined myself to be "Susan's Friend" in this case. This is so, however, only because some rather stringent conditions have been met in this particular case. In particular, it just so happens that the more specific, "bric-a-brac" self-determinations that I've made along the way ("help Susan out when she is in distress," "comfort Susan when she is sad," etc.) could be thought of as "adding up" to a friendship. These bric-a-brac determinations, in other words, wind up coinciding to a significant degree with the practical principles

contained in the extension set of the self-conception "Susan's Friend." Given this overlap—together with the fact that there are no other "stock" self-conceptions for which there is this degree of overlap—makes it acceptable to slide from the bric-a-brac determinations to the claim that I have determined myself to be "Susan's Friend." (Or at the very least, one may wonder if there is any meaningful difference between actually determining oneself to be "Susan's Friend" and merely acting in ways that are consistent with this—and only this—self-conception). Since our linguistic conventions appear to allow for such a shift in these sorts of cases, it is one that I am willing to accept.

As a closing caveat, however, it should be noted that we can allow such "slips" only in very specific circumstances; if we become too liberal on this point, the projectability problem discussed earlier could easily get out of hand. Generally speaking, we should allow a set of bric-a-brac determinations to "add up" to the determination of a more robust self-conception iff (1) the bric-a-brac determinations coincide very closely with the practical principles contained in the relevant "stock" self-conception's extension set, and (2) there are no other stock self-conceptions for which this sort of overlap is possible. If either condition isn't met, then one may begin to wonder why it is acceptable to claim that the agent determined herself to be an "X" in such cases, but it isn't acceptable to claim that she determined herself to be an "A," "B," "C,"...etc. as well (where these other self-conceptions are equally projectable descriptions of the agent given the her bric-a-brac commitments).

7.5 Argument from Self-Knowledge

The third and final argument is for the claim that one can gain more self-knowledge by deliberating in agent-expressive ways. [8] It runs as follows:

---

[8] It is worth noting that the argument from self-knowledge is somewhat "Hegelian" in its inheritance as well, despite the fact that I will discuss it under its own heading. Some writers who have been influenced by Hegel, including Charles Taylor and Neuhouser, discuss the merits of a view roughly along the lines of what I discuss here. While Hegel doesn't discuss the value and process of self-

1. Not only are the extension sets of self-conceptions unfixed (the members of the set can change over time), they are often only vaguely understood by agents when they first acquire these self-conceptions.

2. In the course of performing actions in accordance with the extension set of one's self-conceptions, one typically gains a better understanding of what these extension sets do or should include (and accordingly, a better sense of what the self-conceptions in question amount to).

3. Thus, in the process of acting in accordance with these extension sets (qua extension sets), one comes to a better understanding of what began as relatively inchoate self-conceptions, and accordingly gains an important degree of self-knowledge.

Consider the following example. Upon first deciding that you are a "Citizen of the U.S." in a robust sense (perhaps you make the determination sometime in your teens that being a U.S. citizen means more than just living in the country), you might have a very loose understanding of what this means. You know that a particular government represents you, and that you are both beholden to act in certain ways and are entitled to receive certain benefits from the government, but perhaps not much beyond this. However, in the course of acting as a citizen—e.g. signing up with the selective service, voting in elections, registering a car, having run-ins with the police and the law—you might come to a greater (or at least a different) understanding of what this self-conception means, particularly if you think of yourself under the description "Citizen of the U.S." as you act in these ways. Seeing what your government is doing on your behalf on the nightly news and deliberating about what actions are incumbent upon a citizen when his or her government is pursuing

---

knowledge that I canvass in his discussion of "practical identities," a related point comes up in his discussion of Spirit's progress (and the mechanisms through which it progresses) in *The Phenomenology of Spirit*.

questionable policies forces both renewed consideration of what it means to be a citizen, and ultimately a greater understanding of what one's adoption of this role involves.

This idea is echoed in Neuhouser's discussion of the interpretive aspects of acting in self-expressive ways.

> While social members define their identities in terms of social roles they do not themselves create, executing those roles within the world is less like a mechanical translation of norms into action than an interpretive performance in which individuals determine their actions in accord with their own ideas about how best to exemplify the roles they take as their own (Neuhouser: 108).

The sort of "interpretive performance" that Neuhouser has in mind does indeed seem to be a very important component of what it means both to hold a particular self-conception and to deliberate about how to best express it in one's actions. Of central importance given our aims is the fact that there is a close connection between interpreting what one's acceptance of a given role requires and coming to a clearer understanding of what the role in question involves (and accordingly, a more detailed sense of who one is as an occupier of this role).

Also of interest are Charles Taylor's thoughts on the role that "articulation" plays in interpretive pursuits. "...[A]rticulations are attempts to formulate what is initially inchoate, or confused, or badly formulated. But this kind of formation or reformation does not leave its object unchanged. To give a certain articulation is to shape our sense of what we desire or what we hold important in a certain way" (Taylor (2): 36). Acts of agent-expressive deliberation can help us better articulate our self-conceptions in many cases, which in turn leads to a greater understanding of what it means to hold these self-conceptions.

If one doesn't act in agent-expressive ways, however, the self-knowledge that one gains through one's actions is likely to be more limited. Acting in response to a particular practical demand might give one a better understanding of the principles that one accepts ("so this is what it really means to be compassionate"), and at least potentially a better understanding of oneself as a person committed to this practical demand. But in cases of this sort, what one really learns about (if anything) are the sorts of actions that it is correct to perform. The knowledge gained, in other words, isn't first and foremost *self*-knowledge, but rather knowledge about how one should act more generally (although MCDT tells us that this insight could serve as a form of "indirect self-knowledge" by forming the basis of what could potentially be a better understanding of what one is or should be).

How successful is this argument? A few possible worries are worth considering. There is first a question about when a link between an action and a self-conception must be forged for the agent to gain the sort of self-knowledge in question. It might seem as though one could fail to ever engage in acts of agent-expressive deliberation, and yet nonetheless gain a fair degree of self-knowledge by looking back on past actions and understanding them as expressions of certain self-conceptions.

Unlike the self-determination case, however, it is less clear that acts of retrospective agent-expression can work in the case of self-knowledge. It would seem as though much of the knowledge that one gains in agent-expression cases comes through "interpretive performances" in which one thinks about what sort of action a self-conception demands in the face of unfamiliar circumstances. By thinking through questions of this sort, the agent can in many cases flesh out the details of incompletely understood self-conceptions. The resulting knowledge is a product of performing actions under descriptions that tie them to particular self-conceptions. One cannot, however, alter the descriptions under which certain acts are performed retrospectively.

If I decide to protest the war in Iraq by focusing entirely on the horrors of the lives it will cost (if this is the description under which I act), I cannot at a later date change the description of this act or simply "reinterpret" it as an expression of the self-conception "U.S. Citizen." The act did not result from the realization that it is incumbent upon me as a U.S. Citizen to protest unjust actions performed by my democratically elected government, and so this act cannot be treated retroactively as an "interpretive articulation" of this self-conception.

This is, of course, not to say that I cannot acknowledge at a later point that this sort of action was also incumbent upon me as a "U.S. Citizen." Reflecting again on MCDT, I may come to see later on that this act figures into the appropriateness of holding the self-conception "U.S. Citizen." If the preceding claims about "interpretive performances" are correct, however, it is unlikely that this sort of armchair reflection could lead to the same sorts of insights that I would have gained by struggling with the problem of how a U.S. Citizen ought to act in the face of this particular issue.

Explaining why precisely this is so is not easy, but perhaps the following analogy will help. Suppose that you are curious about what it is like to actually face someone who is threatening to kill you. Perhaps as far as you know, your life has never been in direct jeopardy at the hands of another person. First, it doesn't appear as though you could gain this knowledge by simply imagining this sort of scenario (based on comments by other people, it seems that you would really have to face impending death to understand what it's like). Now suppose that just last week you and a friend came across a man in the street holding a knife out towards you. At the time you thought he was trying to sell you the knife; perhaps he muttered something in French that sounded to you like a business proposition. In a later discussion with your friend, however, you come to learn that the man was actually preparing to stab you. Here too, it doesn't seem as though you can *now* learn what it is like to face

impending death by reconsidering this episode as a case in which you were about to be killed. The reason is simply that at the time of the incident, you understood what was going on under a different description, and your emotional reactions to the incident were based entirely on this description.

A parallel conclusion can be drawn in the case of learning about one's self-conceptions. While reconsidering past actions under different descriptions may yield some insight into what certain self-conceptions are all about, it probably won't yield the degree of self-knowledge that actual agent-expressive acts can. It is by actually facing situations in which you have to decide, for example, what to do as "Susan's Friend" when Susan starts talking about abandoning her career or asks you to lie on her behalf, that this sort of self-knowledge becomes available. It is in the course of reflecting on what true friendship is all about, acting out the conclusions of this deliberation, and assessing your conclusions on the basis of how things turn out (to include Susan's reactions to your choices, etc.) that you come to a better understanding of this role.[9] Reconsidering past acts in which you were true to the principles "advise acquaintances who are undergoing career crises" and "never lie" as cases of agent-expression won't yield the same knowledge. The opportunity for an interactive, interpretive performance has been lost.

A similar line of reply rebuts the (closely related) worry that the relevant self-knowledge might be obtained, not by explicitly expressing one's self-conceptions (either concurrently or retrospectively), but by running through hypothetical situations in one's imagination. The correct response to this thought is to concede that some self-knowledge might be available through this route, but to maintain nonetheless that the potential for gaining self-knowledge in such cases will be limited. The reason is

---

[9] For a very interesting discussion of the ways in which making sense of such cases might require interpretive exchanges and open discussion with other parties effected by one's actions, see Robert Pippin's *Henry James and Modern Moral Life*. See for example (Pippin: 42, 162, 176).

simply that in cases of "virtual agent-expression" you are simulating all of the events, actions, and results of your actions, which might not accurately reflect how things would actually turn out in real situations.

A good illustration of this involves cases in which one thinks ahead about how one ought to act in an upcoming stressful situation, such as a job interview or an impending argument. In such cases, what seemed to be an appropriate course of action in one's imagination often turns out to be clearly inappropriate when one faces the actual situation. For example, before going to a friend's house to discuss what appears to you to be a poor life choice, you might think that a true friend ought to be quite firm in denouncing such a bad idea, since doing so will prevent your friend from harming herself financially. Once you are face to face with your friend having a discussion about this life choice, it might nonetheless become quite clear that a "tough love" stance is absolutely not what a true friend should be adopting in these circumstances. A *successful* agent-expressive interpretive performance may in many cases require an *actual* performance of some sort; hypothetical performance won't always be enough. The upshot is that cases of "virtual agent-expression"—as well as cases in which one attempts to re-describe past actions—might yield some self-knowledge, but not only will this knowledge be quite limited, one risks gaining nothing more than a false sense that one has learned something important about oneself.

7.6 Might There Be Disvalues Associated With Agent-Expressive Deliberation?

To sum up the conclusions reached thus far, we have identified two values that one might have a better shot of achieving if one acts in agent-expressive ways: self-determination (understood as the value of determining who you are going to be), and self-knowledge. At this point, however, it might be asked whether there are any

disvalues associated with acts of agent-expression. Might it be inadvisable, or perhaps irrational, to deliberate in ways that express one's self-conceptions, at least in certain circumstances?

One possible worry is that an overriding concern with figuring out how to act in ways commensurate with one's self-conceptions is a mark of excessive self-concern. In certain cases, it might even seem that an overweening concern with living up to a self-conception would deprive one's actions of their value. If I tend to save drowning people and donate money to environmental causes because doing so is a good way to live up to my own self-image, might these actions have less value than if I performed them without any regard to what they say about me?

Two points are worth mention. First, one could argue that it is isn't irrelevant that the person who acts to express his self-conceptions "Environmental Activist" and "Friend of Greg" may have accepted these self-conceptions in the first place as a response to the value inherent in their associated actions. We could thus say that the value associated with being an "Environmental Activist" or "Friend" is *itself* expressed (although in the weaker, everyday sense of the term) any time one acts in agent-expressive ways with respect to them. For example, our environmental activist may deliberate with an eye to acting in ways appropriate for an "Environmental Activist," but his choice to take up this self-conception and treat it as reason-giving is itself a nod to the value of promoting sound environmental practices. Accordingly, this agent—just like an agent who donates money to the environmental cause simply because he thinks it would be the right thing to do—is acting in response to the perceived value of environmental protection.

Nonetheless, this line of reply doesn't cover all cases. To determine whether this objection is valid in a particular case, we would need to look more closely at the agent's specific motivation. As long as an agent's primary interest is in living up to a

self-conception that he thinks is value-commensurate ("I should donate this money because this is what an environmental activist does, and it truly is good to be an environmental activist"), a charge of egoism seems misplaced. If, however, the agent's main interest is to present the appearance of being a certain sort of person because he thinks he will win favor as a result, or if his real interest is to achieve a certain degree of self-satisfaction, then this objection will have teeth. Acting in agent-expressive ways in and of itself, however, isn't necessarily a bad thing.

Second, note that there is a big difference between worrying about what one can get by performing a given an action, and worrying about *being* a certain sort of person. Korsgaard raises a good point in response to Nagel's charge of egoism when she claims that "...the 'egoism' that my view involves is a matter of being concerned with what you *are*, not a matter of being concerned with what you *get*, and no question of being rewarded is at issue" (Korsgaard (1): 247). In a footnote to this passage, Korsgaard also points out that this is the sort of reply we should offer on behalf of Plato and Aristotle, who are themselves sometimes accused of being overly preoccupied with questions of what one can get out of acting in certain ways. This response might not completely assuage worries that one's "deliberative gaze" should be firmly out upon the world at all times, but it should go a long way toward unseating traditional charges of egoism.

Lastly, one may wonder how often one should engage in acts of agent-expression. Is this a case of "the more, the better," or is it possible there can be too much of a good thing? A survey of how people actually deliberate suggests that one needn't radically change the way one approaches practical questions in response to this chapter's discussion, even if there is some value in agent-expressive action. The simple fact is that few people spend much time engaging in acts of agent-expression (non-agent expressive deliberation seems to be the norm), and yet these people

nonetheless appear to live very fulfilling, value-enriched lives. This suggests that a life dominated by agent-expressive deliberation isn't required by the arguments presented above.

In addition to there being no positive demands on excessive agent-expression, one might worry that too much agent-expressive deliberation could lead to an overly-scripted life, in the sense that one's decisions are "being made for one" by norms inherent in the self-conceptions one accepts. This worry is especially acute in the case of self-conceptions that are determined largely by social forces, like "Brother" or "Patriot" (which, in contrast to self-conceptions like "Friend" are very sensitive to historical accidents, local prejudices, etc.). Perhaps if we deliberate from self-conceptions to actions too often, we wind up living out someone else's view of what a good life ought to look like, rather than making a life of our own.

Where does this leave us? Perhaps the best we can say is that some agent-expressive deliberation is better than none, but perhaps we ought to shoot for only a moderate degree of it. In this respect this chapter's conclusions are fairly modest. MCDT and AAR hold that there is an important relationship between self-conceptions and practical reasons, and the discussion of the past few chapters has shown that we have good reason to attend to the implications of this relationship in our deliberations. Precisely how much attention they are owed is a question for another day.